QUANTITATIVE FINANCE
RESEARCH CENTRE

UNIVERSITY OF
TECHNOLOGY SYDNEY

QUANTITATIVE FINANCE
RESEARCH CENTRE

UTS

THINK.CHANGE.DO

QUANTITATIVE FINANCE RESEARCH CENTRE

Research Paper 159 **May 2005**

Inferential Expectations

Gordon Douglas Menzies and Daniel John Zizzo

ISSN 1441-8010 www.qfrc.uts.edu.au

Inferential Expectations

Gordon Douglas Menzies[1]
University of Technology, Sydney
and Australian National University

Daniel John Zizzo[2]
University of East Anglia

(Version: May 2005)

Abstract

We propose that the formation of beliefs be treated as statistical hypothesis tests, and we label such beliefs *inferential expectations*. If a belief is overturned through the build-up of evidence, agents are assumed to switch to the rational expectation. Rational expectations are shown to be a special (limiting) case of inferential expectations, with the test size α becoming a metric for rationality. When inferential expectations are built into a Dornbusch-style model of the exchange rate, regression tests of Uncovered Interest Parity and the rational expectations version of the term structure both display downward bias in the slope coefficient. We present the results of an experiment that supports inferential expectations.

Keywords: expectations, macroeconomics, rationality, uncovered interest parity, term structure, exchange rate.
JEL Classification Codes: C91, D84, E50, F31.

[1] Tel: +61-2-95147728; fax +61-2-95147711. Email address: gordon.menzies@uts.edu.au
[2] Tel: +44-1603-593668; fax: +44-1603-456259. Email address: d.zizzo@uea.ac.uk
We wish to thank Graham Elliot, Jonathan Kearns, Warwick McKibbin, Adrian Pagan, Malcolm Edey, Donald Hay, Chris Kent, Andrea Patacconi, Tony Richards, Steven Pennings, Joep Sonnemans, Susan Thorp, David Vines, Justin Wolfers, and participants to presentations in Amsterdam, Canberra, Oxford and Sydney for useful advice, Rebecca Chalmers for experimental assistance, Duncan Ford for computer programming, and University of Technology Sydney for financial support. The usual disclaimer applies.

1. Introduction

Rational Expectations (RE) applies the principle of rational behaviour to the acquisition and processing of information and to the formation of expectations (Maddock and Carter, 1982). That is, economic modellers and policy-makers who use RE as a working hypothesis bestow upon their representative agents the ability to calculate mathematical expectations, and, when information is limited, to calculate unbiased and efficient parameter estimators.[3]

This theory has had a central role in macroeconomics since the 1970s. It has been used to model phenomena as diverse as aggregate supply, exchange rates, consumption and economic cycles (Lucas, 1972; Dornbusch, 1976; Hall, 1978; Kydland and Prescott, 1982).[4] Important empirical predictions about exchange rates (Frankel and Rose, 1995) and the term structure of interest rates (Mankiw and Miron, 1986), employ RE as one of their key assumptions.

Despite its influence, the number of alleged empirical failures of RE has built up over the passage of time. In experimental settings, RE predictions are not rejected as null hypotheses in some contexts (see Dwyer et al., 1993), but the most common outcome is that individuals do not hold RE (e.g., Schmalensee, 1976; Blomqvist, 1989; Camerer, 1995; Beckman and Downs, 1997; Swenson, 1997). In addition, experimental research often finds either under-utilization or over-utilization of priors (Camerer, 1995).

One way to meet these criticisms is to build models where agents possess RE in financial markets, while allowing sluggish price adjustment elsewhere. As an approach, this is both simple and tractable, as the ‘overshooting’ exchange rate model of Dornbusch (1976) demonstrated. However, participants in financial markets have been forced to reveal frequent valuations of foreign exchange and securities for decades, and empirical tests for RE have not fared well.

Two tests in particular have proved troublesome. Under the joint hypotheses of RE, risk neutrality and zero transaction costs, the slope coefficients in a regression test of Uncovered Interest Parity (UIP) and the RE version of the term structure should both be unity:[5]

$$\Delta s_{t+1} = \beta\ (i_t - i_t{*}) + u_t \tag{1}$$

$$\frac{\Delta i_{t+1}}{2} = \gamma\ (r_t - i_t) + v_t \tag{2}$$

Typically, the estimated coefficient in the UIP regression, *(1),* is less than unity, and sometimes it is even negative (Frankel and Rose, 1995). Evidence based on *(2)* has not been as damning, but RE remains a seriously contested hypothesis. For example in a study of 3-, 6- and 12-month euro-rates for 17 countries, Gerlach and Smets (1997) found the 76 per cent of regression coefficients were less than unity.

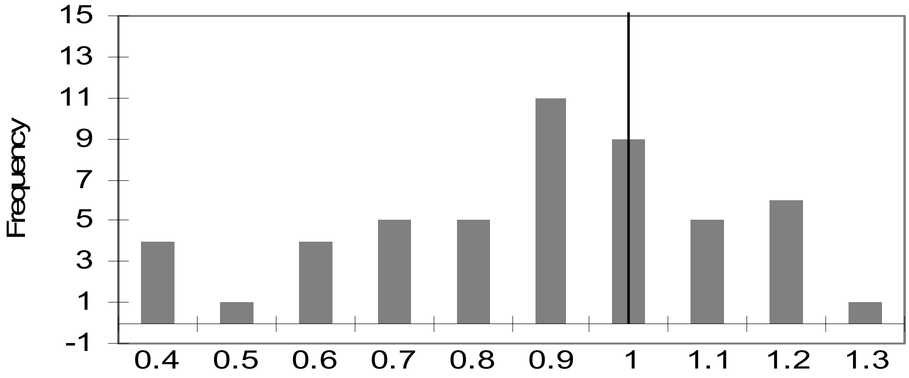
---

[3] The name RE emphasizes the use of mathematical expectations. But any realistic theory of ‘rational’ belief formation must take account of parameter estimation.

[4] RE is also standard in many New Keynesian models (e.g., Woodford, 1991).

[5] The variables *s, i-i** and *r* represent the nominal exchange rate, the short interest differential and the long (here two-period) interest rate. Equation *(1)* can be run with long interest rates, or run in real terms. Both (1) and (2) are sometimes constructed with a constant term.

FIGURE 1. SLOPE COEFFICIENTS FOR RE VERSION OF THE TERM STRUCTURE TESTS (UNDER RE SLOPE=1)



Source: Table 1 in Gerlach and Smets (1997)

Notes: The histogram pools together the slope coefficients of regression of the change in short rates on the interest differential. The regressions have a different form from *(2)* for 3- 6- and 12-month securities, but under RE they should all be unity. Individually, the sample means of the coefficients for each maturity are significantly lower than unity with a maximum p-value of 0.03. The country list comprised European economies, plus Canada, Japan and the US.

One response by macroeconomists is to note that these regressions are testing joint hypotheses. For example, Gerlach and Smets argue that the addition of a time-varying risk premium can rescue RE. Granted the existence of such a premium, they argue along the lines of Mankiw and Miron (1986) that "it is easier to reject the expectations hypothesis in periods in which short-term rates are difficult to predict" (Gerlach and Smets, 1997, pg. 306).[6] Attempts to explain the failure of UIP using a time-varying risk premium have run up against survey evidence which speaks strongly against RE (see Frankel and Froot, 1987).

In this paper we propose a simple alternative to RE which is consistent with the above regression results, and for which we can offer some experimental evidence. While we cannot rule out the possibility of time-varying risk premia, our approach is to explain the failure of these regressions using the notion of *belief conservatism*.

Formally, we suggest that belief formation be treated as a Neyman-Pearson hypothesis test, dubbed *inferential expectations* (IE). We assume that when a belief is overturned agents switch to RE. Thus, RE is thus a special case of IE if agents are unconcerned about mistakenly changing their beliefs (the test size α equals unity), or if there is so much information available about a parameter that it is known with certainty (the sampling distribution of the estimator collapses to a point at the limit) leading to the rejection of any incorrect null.

The intuition of IE is that economic agents hold beliefs that are subject to falsification by new information, in much the same way that they are in conventional statistical hypothesis testing.

[6] However, Kool and Thornton (2004) raise the importance of outliers for Mankiw and Miron's conclusions.

In particular, a change in beliefs requires new information that exceeds a threshold, modelled here by statistical significance. Neyman-Pearson hypothesis tests can be postulated to operate as 'fast and frugal' heuristics (in the style of Gigerenzer et al., 1999) held by near-rational agents.

Thus we assert that beliefs about economic variables tend to be more subject to periods of inertia interspersed with occasional discrete shifts than what would be implied by RE. A key implication is that agents do not fully process new information every period - possibly because they underweight its value and possibly because they do not pay enough conscious attention to it -, even if they fully use information occasionally. IE therefore has much in common with other models of sluggish belief adjustment (Mankiw and Reis, 2002, 2003).

The alleged reluctance to change one's mind despite some contrary evidence is not a new idea in the philosophy of science (Kuhn, 1970), and it has recently gained some acceptance in behavioural macroeconomic where "herding and *procrastination* help explain the significant departures of real-world economies from the competitive general-equilibrium model" (Akerlof, 2002, pg. 428, our italics). Indeed, IE is built into the methodology of any social science that uses classical hypotheses tests as decisive evidence for changing beliefs. Whenever an academic economist uses a test to convince a colleague about the truth of a proposition, she implies that the belief in question will *not* be altered until probability values such as 0.05 or 0.01 are attained.[7] Thus, IE potentially explains temporary deviations from strict rationality while still imposing a plausible and simple structure on expectations which is consistent with economists' own practices in statistical data analysis.[8]

We also present the results from a simple experimental environment where IE outperforms RE. Agents are asked to repeatedly declare the probability that a randomly chosen urn is of a particular kind, based on successive draws of balls from the chosen urn. An observed phenomenon that is explicable under IE (but not RE) is that agents sometimes do not change their declared probabilities in spite of receiving new information.

Our work is perhaps closest in spirit to Goldberg and Frydman (1996) and Frydman and Goldberg (2003), who allow agents to conduct hypothesis tests over models.[9] Their research program, tracing its roots back at least to an informal discussion by Rappaport (1985), allows for departures from rational beliefs, though it is more radical than the temporary departures envisaged in this paper.

The paper is organized as follows. In section 2 we explain the theory underlying IE and describe a Dornbusch-style model with IE. In section 3 we prove that regressions *(1)* and *(2)* will have downward-biased coefficients. Section 4 provides the experimental evidence for IE. Section 5 concludes.

---

[7] Hypothesis tests are used despite the availability of an alternative approach based on Bayesian inference (e.g., Zellner, 1988). We do not wish to imply that economic theorizing relies exclusively, or even primarily, on statistical hypothesis tests. Nevertheless, one author received criticism from a referee in a top-tier journal because significance was accorded to a test with a p-value of 0.052.

[8] It is also consistent with modelling practice in the Markov Switching literature, where passing a threshold, however tentatively, leads to a new regime (Hamilton, 1989).

[9] Our work can also be seen as related to Foster and Peyton Young's (2003) game-theoretical work on hypothesis testing by bounded-rational agents on their opponents' repeated games strategies.

2. Inferential Expectations and Exchange Rate Determination

In all IE models, there is a *cognitive target*. This is a set of state variables or parameters that are believed to be in one of two states, described by the null hypothesis H_0 and the alternative hypothesis H_1.[10] There is also a *signal* which is a model variable that provides information about the cognitive target. Finally there is a *test statistic* and a *rejection region* that are defined conventionally.

Formally, let x be a vector of parameters or random variables that are part of a data-generating process for a random variable p. Granted some economic significance to x, agents form beliefs about it, based on n stochastic signals p_i (for $i = 1, \ldots, n$).

The rational expectation is the mathematically best guess for x. The inferential expectation is the mathematically best guess for x, subject to conservatism about changing beliefs (made operational by a Neyman-Pearson hypothesis test of size α), and incorporating any testing shortcuts that qualify as a 'fast and frugal' heuristic. When the concern about changing beliefs becomes vanishingly small ($\alpha \rightarrow 1$), IE and RE coincide.[11] The *cognitive target* is x, the *signal* p_i, the *test statistic* some function of the p_i's, and the *rejection region* are the values of the test statistic that lead to a rejection of a Neyman-Pearson hypothesis test of size α.

Under the assumption of IE, there is first 'under-use' of marginal pieces of information (by comparison with RE) and then 'over-use' (when beliefs change). As indicated in the introduction, RE is nested in IE when:

1. α equals unity. This is clear from the fact that α equals unity implies a rejection for any value of the test statistic. That being so, the RE belief is constantly embraced.

2. x is a single parameter and n equals infinity (and memory is unbounded). If the p_i's are combined in an estimator for x whose sampling distribution variance is decreasing in n, the estimator collapses to a single point at the true value. All nulls will be overturned (except if the null is correct), and RE beliefs embraced.

To provide a macroeconomic context, consider a model in the spirit of Dornbusch (1976):

$$m - p_t = \bar{y} - B_i i_t \tag{3}$$

$$p_{t+1} - p_t = -B_p(p_t - p_{ss}). \qquad 0 < B_p < 1 \tag{4}$$

$$E[s_{t+1}] - s_t = i_t - i^* \quad \Leftrightarrow \quad s_t = -(i_t - i^*) + E[s_{t+1}] \tag{5}$$

$$r_t = \frac{1}{2}(i_t + E[i_{t+1}]) \quad \Leftrightarrow \quad \frac{1}{2}(E[i_{t+1}] - i_t) = r_t - i_t \tag{6}$$

[10] In many applications, it will be a single variable or parameter.

[11] This is true regardless of any shortcuts used in the testing procedure. If the test size is unity, then hypothesis testing is suspended (along with any shortcuts about distributional assumptions, etc.) because a Neyman-Pearson hypothesis test minimizes the size of the probability of a type II error given the test size (probability of a type I error). That is, given a test size of unity, the best way to minimize the chance of falsely believing the null is to *always* reject it.

All variables are in logs, except the nominal short and long (two-period) interest rate, respectively i_t and r_t. All parameters are positive. Time zero is divided into pre-money-shock 0^- and post-money-shock 0^+; afterwards, t = 1, 2, 3,……∞. The exchange rate, money, prices (domestic and foreign), output and foreign interest rates are *s, m, p, p*,* $\bar{y}$ and *i**. The latter three variables, and m_{0^-}, are normalized to zero. Two steady states occur; one prior to a money shock (at 0^-) and the one after the shock has completely dissipated (at ∞). A monetary contraction Δm (<0) occurs at time 0^+ and is sustained forever. The pre-shock *m* of 0 implies a steady state of $p_{0^-} = 0$ and $s_{0^-} = 0$ (the latter from purchasing power parity). Equation *(3)* can be either an LM curve or a quasi-Taylor rule. For the latter, it is re-expressed with *i* as the subject and *m* as the nominal income target.[12] The standard Dornbusch assumption of sticky prices is adopted.

The standard Dornbusch solution assumes RE. That is, $E(s_{t+1})=s_{t+1}$ and $E(i_{t+1})=i_{t+1}$. The eigenvalue of the system is $(1\text{-}B_p)$, and the solution from $t=0^+$ is given below (derivations in Appendix A).

$$i_{t+1} = (1-B_p)i_t \Rightarrow$$

$$p_t = -\Delta m(1-B_p)^t + \Delta m \tag{7}$$

$$i_t = \frac{-\Delta m}{B_i}(1-B_p)^t \tag{8}$$

$$s_t = \frac{\Delta m}{B_i B_p}(1-B_p)^t + s_\infty = \frac{\Delta m}{B_i B_p}(1-B_p)^t + \Delta m \tag{9}$$

The IE solution requires a plausible null belief. We assume that this is the belief that the central bank is engaging in a managed float, and that the initial steady-state exchange rate ($s_{0^-} = 0$) is its target level of the currency.[13] Until enough evidence builds up, agents believe that next period the central bank will defend the currency by driving the log-exchange rate back to zero, and keep it there forever.[14] Appendix A shows how this outcome could be achieved by driving *i* to zero next period, and keeping it there forever. To parameterize the IE hypothesis test, we assume that agents are making inference about the future values of *s* based on the latest time series observation on *p*.[15]

Formally, the inference procedure has to decide between:

H_0: $s_{t+j} = 0, j > 0$ (implying $i_{t+j} = 0\ j > 0$).
H_1: A standard Dornbusch money contraction will be sustained; $m_{t+j}=\Delta m\ j > 0$.

---

[12] The model has a block recursive structure, which we will exploit to obtain analytic solutions under IE. A money shock drives *p* and *i*, via (long run) purchasing power parity, the quasi-Taylor rule, and the partial adjustment process for prices. Then, *s* and *r* are determined by *(5)* and *(6)*. Purchasing power parity implies $s+p^*-p = 0$.
[13] This target can only be met by setting the appropriate level of interest rates, and departures from the target are possible. Therefore, the institutional setup of the model is a managed float (rather than a fixed rate regime) where the instrument is the interest rate (rather than foreign exchange intervention).
[14] Or naive agents could view recent moves in *i* and *s* as ‘turbulence’, which will vanish next period.
[15] Other signals are possible. Agents could stop believing the bank when non-zero interest rates persist.

Interpreting *(3)* as a quasi-Taylor rule, H_0 implies the central bank is permitting a temporary change in the price level,[16] while H_1 implies a permanent accommodation to a changed price level. In the latter case, the solution to the model is given by *(7)* to *(9)*. Under H_1, the steady state level of p and e would both be Δm, from $t=\infty$ in *(7)* and *(9)*, consistently with purchasing power parity.

One possible interpretation of this analysis is that IE tells a story of the central bank having to give away a currency target, when the credibility strain becomes too great.[17] In this interpretation, the shock to m represents the implementation of policy that puts long-run purchasing power parity and the central bank exchange rate target at odds with one another.

In the Dornbusch model, this would be recognized immediately, and the central bank would be ignored. RE agents work out the implied price level from the shock, and use purchasing power parity to calculate the implied long-run exchange rate. They then solve backwards using forecast interest differentials, forcing the current exchange rate to 'jump' once to its 'overshooting' value. In IE, the story is more subtle: IE agents *initially believe that the central bank will defend the currency* from next period. To solve for the exchange rate, they note that the long-run exchange rate will be zero and that all future interest differentials must be too (see Appendix A). Therefore, the current exchange rate does not have the long-run shock built in, nor does it account for future domestic returns. Initially it 'jumps', but only by the magnitude of the *current* (non-zero) interest differential (see *(11)*), since this is the only influence that IE agents recognize. Later on, when agents become disillusioned with the central bank, they factor in future returns and the shock to the long run, and the exchange rate jumps again, this time to attain the RE solution.

A rejection region and test statistic that is consistent with our parameterization of IE would be:

$$\text{Reject } \mathrm{H}_0 \text{ when } \mathrm{P}_t \le \alpha\, \mathrm{P}_{0^-} + (1-\alpha)\,\mathrm{P}_{\infty,H1} = \alpha\ 0 + (1-\alpha)\Delta M = (1-\alpha)\Delta M. \qquad (10)$$

where α is the test size.[18] If $\alpha = 0$, the null will never be overturned, as would be the case in any hypothesis test. If $\alpha = 1$ the null will never be believed, and the model reverts to the RE solution, as required by IE.[19] At a 0.05 significance level agents believe that s and i will revert to their initial values (0) in subsequent periods, until prices have changed by 95 per cent of what is implied under H_1.

---

[16] H_0 implies steady-state s is zero, purchasing power parity therefore implies that steady-state p is likewise zero.

[17] In our stylized shock, the central bank is fighting against an appreciation, while it pursues a deflationary policy. Naturally, it is possible to reverse the shock, so that the central bank is fighting against a depreciation while pursuing inflationary policy. However, to avoid negative nominal interest rates (a problem generally ignored in Dornbusch formulations) the $i^*=0$ normalization would have to be dropped, creating less transparent algebra.

[18] More generally, the weight in the weighted-average would be $\phi(\alpha)$ and $1-\phi(\alpha)$, where $\phi(\alpha)$ has the property that $\phi(1) = 1$ and $\phi(0) = 0$.

[19] The second case where IE becomes RE (infinite sample size and unbounded memory) is not relevant if the test is based only on the most recent p (n is unity), as here. It would be possible to define a test statistic using a combination of the p's (n>1) such that the sampling distribution variance was decreasing in n. Time intervals could also be defined over progressively smaller increments, so that as $n\to\infty$ the null would be rejected immediately, ushering in the second case in the limit.

Thus, the *cognitive target* in this model is the set of all future exchange rates, the *signal* is p, the *test statistic* at time t is the most recent price level p_t and the *rejection region* is given by *(10)*.

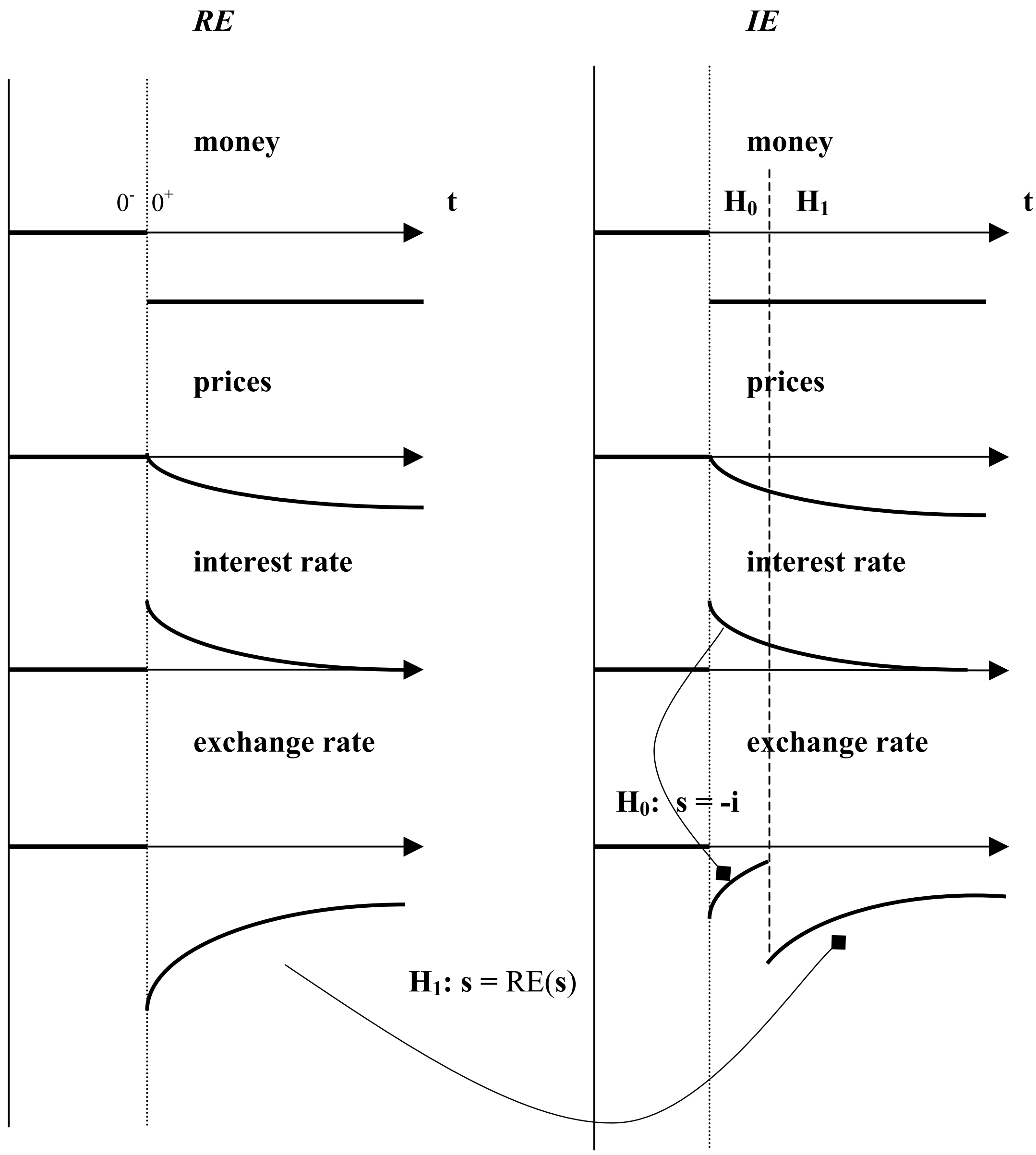
It remains for us to specify the time path for s and r under H_0, which alters the model by violating $E(s_{t+1})=s_{t+1}$ and $E(i_{t+1})=i_{t+1}$. Under H_0, s and i are forecast to revert to the original steady state next period, so that $E(s_{t+1})=E(i_{t+1})=0$. That being so, the actual evolution of i and p – which are not affected by expectations – are as before, while the solutions for s and r until H_0 is rejected come from (5) and (6) with $E(s_{t+1})=E(i_{t+1})=0$.

$$s_t = -i_t = \frac{\Delta m}{B_i}(1-B_p)^t \qquad \textit{prior to rejection of } H_0 \tag{11}$$

$$r_t = \frac{i_t}{2} = \frac{-\Delta m}{2B_i}(1-B_p)^t \qquad \textit{prior to rejection of } H_0 \tag{12}$$

When H_0 is rejected, the exchange rate jumps to the Dornbusch path. This can be illustrated with the standard diagram.

FIGURE 2. DORNBUSCH SETUP WITH RE AND IE
(Under H_1 Left & Right Panels Identical)



Notes: The left-hand panel shows the standard dynamics from Dornbusch (1976) for a monetary contraction. The right-hand panel show the IE dynamics. Prior to H_0 being overturned, agents in the foreign exchange market ignore future (positive) interest rate differentials, and the appreciated (lower) transversality condition for the exchange rate. When H_0 is discarded for H_1 (as the shaded region is entered) the exchange rate jumps to the RE solution.

Importantly, IE has the capacity to deliver a sudden, and potentially large, change in a model variable – here the exchange rate – for a small increment of information, at the instant H_0 is rejected.

We now turn to two implications of IE in this model.

3 Downward Bias in OLS Regressions

We now show that data on *s* and *r* generated by this model with IE will, when placed in the OLS regressions *(1)* and *(2),* create downward bias in the parameters.

Theorem 1: If H_0 is believed for at least one period, the OLS coefficient from a regression of Δs_{t+1} on i_t will be less than unity.

Proof: Let t^* be the time period in which H_0 is overturned. The numerator of the OLS coefficient can be decomposed into three terms; the cross product summed prior to, at, and after t^*-1.

$$\hat{\beta} = \frac{\sum_{t=0}^{n} \Delta s_{t+1} i_t}{\sum_{t=0}^{n} i_t^{\ 2}} = \frac{\sum_{t=0}^{t^*-2} \Delta s_{t+1} i_t + \Delta s_{t^*} i_{t^*-1} + \sum_{t=t^*}^{n} \Delta s_{t+1} i_t}{\sum_{t=0}^{n} i_t^{\ 2}} \tag{13}$$

We evaluate each quantity in the numerator separately.
Taking the first term, *(11)* implies that the change in s_t is:

$$\Delta s_{t+1} = -(i_{t+1} - i_t) = -([1 - B_p] i_t - i_t) = B_p i_t \tag{14}$$

$$\therefore \quad \sum_{t=0}^{t^*-2} \Delta s_{t+1} i_t = \sum_{t=0}^{t^*-2} B_p i_t^{\ 2}$$

Parenthetically, this establishes Theorem 1 for the special case of $\alpha = 0$. Agents never switch from H_0 and an OLS regression of Δs on *i* will fit perfectly, with a coefficient B_p, which is assumed less than unity (see *(4)*).

The middle term of the numerator in *(13)* involves a change in the exchange rate as it jumps from the H_0 path at time t*-1 to the H_1 path at time t*. (See the bottom of right-hand panel in Figure 2.) The H_1 value of the exchange rate is given by a combination of *(8)* and *(9)*, and reflects a realization that the terminal value for the exchange rate has jumped down by Δm.

$$\begin{aligned}
\Delta s_{t*} &= s_{t*} - s_{t*-1} \\
&= \left(-\frac{i_{t*}}{B_p} + \Delta m \right) - (-i_{t*-1}) \\
&= \left(-\frac{(1-B_p)i_{t*-1}}{B_p} + \Delta m \right) + i_{t*-1} \\
&= \Delta m + (2 - \frac{1}{B_p}) i_{t*-1}
\end{aligned}$$

$$\therefore \Delta s_{t*} i_{t*-1} = (\frac{\Delta m}{i_{t*-1}} + 2 - \frac{1}{B_p}) i^2{}_{t*-1} \qquad (15)$$

Since B_p is no larger than unity, $2\text{-}1/B_p$ can never exceed unity. Since Δm is negative, the bracketed term on the RHS must always be less than unity.
The last term on the numerator of *(13)* is evaluated under H_1 so that *(5)* holds, with $E(s_{t+1})=s_{t+1}$ (since H_1 is embraced) and $i^*=0$. Clearly:

$$\sum_{t=t*}^{n} \Delta s_{t+1} i_t = \sum_{t=t*}^{n} i_t{}^2 .$$

Collecting the three parts, the OLS numerator is a weighted average of i^2, with weights prior to the observations on the H_1 path strictly less than unity.

$$\hat{\beta} = \sum_{t=0}^{n} w_t i_t{}^2 \Big/ \sum_{t=0}^{n} i_t{}^2 \qquad w_t = \begin{cases} B_p & t < t*-1 \\ \dfrac{\Delta m}{i_{t*-1}} + 2 - \dfrac{1}{B_p} & t = t*-1 \\ 1 & t > t*-1 \end{cases} \qquad (16)$$

Apart from the trivial case where H_0 is never believed (w_t=1 always), the OLS coefficient must be strictly less than unity. Q.E.D.

Lemma 1: One of the weights in the OLS numerator can be negative.

Proof: For small values of B_p, noting that

$$i_t = -(\Delta m / B_i)(1-B_p)^t \Rightarrow \Delta m / i_{t-1} = -B_i / (1-B_p)^{t-1}$$

the weight applied to $i^2{}_{t*-1}$ on the numerator of the OLS coefficient could be *negative*. This raises the possibility that the numerator, and therefore the whole estimate, could be negative. Q.E.D.

The intuition of Theorem 1 has already been outlined. Agents are not factoring in the altered steady-state purchasing power parity value of the exchange rate, which restrains the appreciation, compared with the RE solution under H_1. Furthermore, from *(14)*, the future change in the exchange rate is muted by a factor B_p compared with the H_1 solution, reflecting a failure to factor in the future returns available from holding the domestic currency. In other words, the coefficient is biased down partly because the H_0 expectation of interest rates (zero for every future period) makes agents too pessimistic about returns, and partly because the jump appreciation (fall in *s*) between the H_0 and H_1 exchange rate paths occurs while the actual interest differential is positive. If the latter jump is large enough, it can make the coefficient negative, as implied by Lemma 1.

We conclude the section on UIP by noting that the regression *(1)* is often run with a constant term. Simulations confirm that the slope coefficient remains sensitive to the deep parameter B_p, and that a negative $\hat{\beta}$ obtains when B_p is low. High values of α tend to magnify *(15)*, making the coefficient negative, as the jump between paths occurs when i_{t^*-1} is high.[20]

TABLE 1– UNCOVERED INTEREST RATE REGRESSIONS WITH CONSTANT: MEAN SLOPE COEFFICIENT VALUES CLASSIFIED BY α AND β_p

	α									
β_p	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.196	0.166	0.087	-0.042	-0.23	-0.435	-0.74	-0.946	-1.198	-1.885
	(0.111)	(0.171)	(0.27)	(0.38)	(0.497)	(0.596)	(0.715)	(0.783)	(0.856)	(1.019)
0.2	0.225	0.204	0.148	0.061	-0.014	-0.125	-0.29	-0.538	-0.913	-0.913
	(0.089)	(0.112)	(0.157)	(0.208)	(0.243)	(0.285)	(0.337)	(0.397)	(0.465)	(0.465)
0.3	0.314	0.289	0.224	0.156	0.037	-0.179	-0.179	-0.578	-0.578	-0.578
	(0.086)	(0.101)	(0.132)	(0.157)	(0.191)	(0.235)	(0.235)	(0.285)	(0.285)	(0.285)
0.4	0.399	0.375	0.331	0.244	0.06	0.06	-0.351	-0.351	-0.351	-0.351
	(0.09)	(0.098)	(0.11)	(0.13)	(0.159)	(0.159)	(0.187)	(0.187)	(0.187)	(0.187)
0.5	0.5	0.471	0.406	0.25	0.25	-0.165	-0.165	-0.165	-0.165	-0.165
	(0.089)	(0.094)	(0.104)	(0.12)	(0.12)	(0.125)	(0.125)	(0.125)	(0.125)	(0.125)
0.6	0.595	0.547	0.415	0.415	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	(0.09)	(0.094)	(0.101)	(0.101)	(0.081)	(0.081)	(0.081)	(0.081)	(0.081)	(0.081)
0.7	0.678	0.568	0.568	0.143	0.143	0.143	0.143	0.143	0.143	0.143
	(0.092)	(0.094)	(0.094)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
0.8	0.718	0.718	0.274	0.274	0.274	0.274	0.274	0.274	0.274	0.274
	(0.092)	(0.092)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
0.9	0.867	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393	0.393
	(0.093)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)

Notes: The table reports mean slope coefficients (and mean standard errors in parenthesis) across simulations with different values of β_i.

We now turn to *(2)*, the term structure regression.

Theorem 2: If H_0 is believed for at least one period, the OLS coefficient from a regression of Δi_{t+1} on $r_t - i_t$ will be biased downwards (strictly less than unity).

Proof: Under H_0:

$$r_t = \frac{1}{2} i_t \Rightarrow r_t - i_t = -\frac{1}{2} i_t$$

$$i_{t+1} - i_t = (1 - B_p) i_t - i_t = -B_p i_t$$

$$\therefore \quad \frac{(i_{t+1} - i_t)}{2}(r_t - i_t) = \frac{B_p i_t^2}{4} \qquad \textit{and} \qquad (r_t - i_t)^2 = \frac{i_t^2}{4}.$$

[20] Appendix B describes the simulations.

Under H_1, equation *(6)* holds (with $E(i_{t+1})=i_{t+1}$) and so:

$$r_t = \frac{1}{2}(i_t + i_{t+1}) \Rightarrow \qquad r_t - i_t = \frac{1}{2}(i_{t+1} - i_t)$$
$$= \frac{1}{2}((1-B_p)i_t - i_t) = -\frac{B_p}{2} i_t$$
$$\therefore \qquad \frac{(i_{t+1} - i_t)}{2}(r_t - i_t) = \frac{B_p{}^2 i_t{}^2}{4} \qquad and \qquad (r_t - i_t)^2 = \frac{B_p{}^2 i_t{}^2}{4}.$$

The OLS estimator is relatively simple to calculate, because the only variable involving a future change *(Δi_{t+1})* is unaffected by expectations, and so does not involve a complex jump between paths at the switch between H_0 and H_1 (as *s* did in *(1)*). Thus we can partition the sums into the terms where the respective hypotheses hold.

$$\hat{\gamma} = \frac{\sum_{t=0}^{n} \frac{\Delta i_{t+1}}{2}(r_t - i_t)}{\sum_{t=0}^{n}(r_t - i_t)^2} = \frac{\sum_{H_0} \frac{\Delta i_{t+1}}{2}(r_t - i_t) + \sum_{H_1} \frac{\Delta i_{t+1}}{2}(r_t - i_t)}{\sum_{H_0}(r_t - i_t)^2 + \sum_{H_1}(r_t - i_t)^2}$$

$$= \frac{\sum_{H_0} \frac{B_p i_t{}^2}{4} + \sum_{H_1} \frac{B_p{}^2 i_t{}^2}{4}}{\sum_{H_0} \frac{i_t{}^2}{4} + \sum_{H_1} \frac{B_p{}^2 i_t{}^2}{4}} = \frac{B_p \sum_{H_0} i_t{}^2 + \sum_{H_1} B_p{}^2 i_t{}^2}{\sum_{H_0} i_t{}^2 + \sum_{H_1} B_p{}^2 i_t{}^2} < 1 \quad Q.E.D. \qquad (17)$$

Clearly, a negative $\hat{\gamma}$ is impossible. As was the case for the UIP regression, *(2)* is often run with a constant term. As reviewed in Appendix B, simulations confirm that the slope coefficient remains sensitive to the deep parameter B_p but the value of the slope coefficient is relatively robust to different values of α.

TABLE 2: TERM STRUCTURE REGRESSIONS WITH CONSTANT:
MEAN SLOPE COEFFICIENT VALUES CLASSIFIED BY α AND β_p

	α									
β_p	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.099	0.094	0.088	0.084	0.081	0.080	0.079	0.079	0.080	0.091
	(0)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.006)	(0.007)	(0.008)	(0.016)
0.2	0.197	0.195	0.192	0.190	0.190	0.190	0.191	0.197	0.220	0.220
	(0.001)	(0.002)	(0.003)	(0.005)	(0.006)	(0.008)	(0.011)	(0.015)	(0.022)	(0.022)
0.3	0.298	0.297	0.295	0.295	0.297	0.304	0.304	0.334	0.334	0.334
	(0.001)	(0.002)	(0.004)	(0.006)	(0.009)	(0.014)	(0.014)	(0.023)	(0.023)	(0.023)
0.4	0.399	0.398	0.398	0.398	0.405	0.405	0.435	0.435	0.435	0.435
	(0.001)	(0.002)	(0.004)	(0.006)	(0.011)	(0.011)	(0.02)	(0.02)	(0.02)	(0.02)
0.5	0.499	0.499	0.499	0.503	0.503	0.529	0.529	0.529	0.529	0.529
	(0.001)	(0.002)	(0.004)	(0.008)	(0.008)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
0.6	0.600	0.599	0.601	0.601	0.620	0.620	0.620	0.620	0.620	0.620
	(0.001)	(0.002)	(0.005)	(0.005)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
0.7	0.700	0.700	0.700	0.711	0.711	0.711	0.711	0.711	0.711	0.711
	(0.001)	(0.002)	(0.002)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
0.8	0.800	0.800	0.804	0.804	0.804	0.804	0.804	0.804	0.804	0.804
	(0.001)	(0.001)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
0.9	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900	0.900
	(0)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Notes: The table reports mean slope coefficients (and mean standard errors in parenthesis) across simulations with different values of β_i.

We conclude this section by noting that our model connects interest-rate forecast errors and coefficient bias. Under the null, agents make errors in forecasting future interest differentials, since they incorrectly believe them to be zero. Thus, the downward bias in the expectation hypothesis and UIP coefficients is consistent with agents making interest-rate forecast errors (Mankiw and Miron, 1986, and Gourinchas and Tornell, 2004).[21]

5 Inferential Expectations and Experimental Evidence

5.1 *Introduction*

In this section we briefly describe an individual choice experiment designed to test whether IE has significantly greater explanatory power than RE or, to put it differently, whether there are subjects for which assuming that $\alpha < 1$ provides a better fit.[22] The individual choice design is best suited to test the idea of IE in its cleanest form, i.e. without having to worry about the strategic considerations that would arise from a strategic or market setting. As discussed in the introduction, the design is also meant to capture beliefs conservatism in a simple way: we predict that, contrary to RE, agents sometimes do not change beliefs in response to new information.

There were two urns reflecting two possible states of the worlds, namely different combinations of white and orange balls. The true state of the world was chosen randomly, and subjects received signals about its nature by the means of random ball draws with replacement from the 'chosen urn'. The prior probability of an urn being chosen was 0.5 at the start of the experiment, but should have then evolved differently according to the observed sequence of white and orange balls being drawn and, importantly, according to different models of expectation formation. We next describe the experimental design in more detail, and then move to the experimental predictions and results.

5.2 *Experimental Design*

The experiment was run in at the School of Finance and Economics, University of Technology Sydney, in September 2003.[23] Recruitment was through lecture announcements, posters, and UTS Online (a local forum for electronic notices). Recruits were predominantly, though not exclusively, undergraduate students. There were six experimental sessions, three for each of the two experimental conditions; all sessions had six subjects except the last one, which had seven, for a total of 37 subjects. The experiment lasted about two hours, and paid an average of 31.42 Australian dollars ($A).[24] The experiment was in two stages, structurally

---

[21] We leave the strength and direction of this relationship for future research, depending as it does on the choice of a fairly stylized null and alternative hypothesis. For the regression without a constant term (17), the slope is increasing in α. A higher test size reduces the time H_0 prevails and therefore reduces the interest-rate sums over H_0. Thus, a better interest-rate forecasting performance (a higher α) increases the coefficient (consistently with Mankiw and Miron, 1986). Appendix A proves that the IE profile of interest rate expectations – namely $E(i_{t+1})=0$ up until the null is overturned – can be consistent with the estimated persistence of interest rate predictions for reasonable parameters (as in Gourinchas and Tornell, 2004).

[22] For space constraints, we focus only on IE as alternative to RE in this paper. Menzies and Zizzo (2003) also consider two adaptive expectations algorithms, and we plan to consider one or two other algorithms in additional research. None of the algorithms we are aware of, however, can explain the key 'no switch' finding discussed in this section.

[23] It was approved by the UTS Ethics Committee. The experimental instructions can be found in Appendix C.

[24] This was roughly equal to 25 US dollars.

unrelated to one another; in this paper we focus only on the first stage, which had six periods of fifteen rounds each and took over 75% of the session time.

At the start of the session subjects faced a table on the top of which there were two identical urns, a set of white and orange balls in a basket, and a screen. In the 0.7 condition, the experimenter (a) showed subjects that both urns were empty, (b) in front of the subjects, he took seven white balls and three orange balls and placed them in one of the two urns (Urn 1 in what follows) (c) and he took three white balls and seven orange balls and placed them in the other urn (Urn 2 in what follows); (d) he then hid both urns behind the screen. The 0.6 condition was identical to the 0.7 condition, except that Urn 1 got six white balls and four orange balls, and Urn 2 got four white balls and six orange balls.

At the start of each period subjects were reminded about the period number and then one of the two urns was randomly chosen by the flip of a coin in front of the subjects, and put on display. Let us label this urn the 'chosen urn'. It was made clear to the subjects that the probability of Urn 1 being chosen was 50% at the start of each period, but they were not told which urn had actually been chosen.

At the start of each round the experimenter drew a ball from the chosen urn, showed it to the subjects and then put it back in; subjects were asked to write down the ball color in correspondence to the correct period and round in their answer booklet, and then had to make a probability *guess*, between 0% and 100%, on how likely it was that the chosen urn was Urn 1. Subjects were told not to change choices made in previous rounds.[25]

Once a period was completed, the following period got started with a new flip of the coin, up to the end of the 6th period. It was made clear to the subjects that the probability an urn was chosen was entirely independent of the probability that it had been chosen in previous periods.[26]

Payment was based on the guess made in a randomly chosen period and round picked at the end of the experiment. A standard quadratic scoring rule (e.g., Davis and Holt, 1993) was used in relation to this round to penalise incorrect answers: if the chosen urn was Urn 1, then subjects got $25 - 25 \times (guess - 1)^2$ \$A; if the chosen urn was Urn 2, then subject got $25 - 25 \times guess^2$ \$A. Subjects were provided with a payment table detailing the payment for each level of error, without need of any computation on their part (see Appendix C). There was also a participation fee of 8 \$A.

5.3 *Experimental Predictions*

Rational Expectations. The prior probability was set at 0.5. As information flowed in, RE (or, equivalently, IE with $\alpha = 1$) predicted straightforward Bayesian updating depending on whether white or orange balls were drawn.

Inferential Expectations. The IE signal, say s_i, is the drawn ball so that $s_i = 1$ for a white ball and 0 for an orange ball.[27] The cognitive target is the probability that the chosen urn is Urn 1,

---

[25] We shall return to this point towards the end of section 5.3.

[26] A questionnaire administered to the subjects at the start of the experiment ensured that this, and other key points, were clear. The experimenters gave clarifications to the subjects who got answers wrong on the questionnaire.

[27] Naturally, this terminology is unrelated to the Dornbusch model, where *s* was the exchange rate.

or, relatedly, the total probability of drawing a white ball, given the beliefs about the chosen urn, denoted p_w.[28] It is simpler to describe the test with the second cognitive target (the total probability of a white ball), since the test statistic becomes the sample proportion of ones, $\hat{s}_i$. With both urns equally likely at the start, the original null is that the total probability of a white ball is 0.5.[29] The appropriate test here is a two-sided test, and we assume that agents use one of two alternative shortcuts to determine the rejection region. The first method, the Normal approximation method,[30] requires agents to maintain the belief corresponding to the null hypothesis until signal i is received such that:

$$\frac{\left|\hat{s}_i - p_w\right|}{\sqrt{\frac{p_w(1-p_w)}{i}}} > \left|z_\alpha\right| \quad where \;\; p_w = P(white \mid Urn1)P(Urn1) + P(white \mid Urn2)(1 - P(Urn1))$$

The second method is nonparametric and relies on Chebyshev's inequality[31]:

$$P(|\, z \,|> k) \leq \frac{1}{k^2}$$

where z is a standardized random variable (the distance from the mean in units of standard deviations), and the weak inequality is relevant for a discrete random variable. If the probability of getting an observation more than k standard deviations away from the mean is less than $1/k^2$, we may set $1/k^2$ equal to α, and make a rare event statement.[32]

The main advantage of employing Chebyshev's inequality is that it only requires the computation of mean and variance, sidestepping the need for distributional assumptions, albeit at a loss of statistical power. If Chebyshev's inequality is used, we reject H_0 if:

$$\frac{\left|\hat{s}_i - p_w\right|}{\sqrt{\frac{p_w(1-p_w)}{i}}} > \frac{1}{\sqrt{\alpha}}.$$

Both shortcuts can be considered as consistent with a view of IE as a fast and frugal heuristic. In what follows we label IE_N the predictions of IE complemented with Normal approximation and IE_C the predictions of IE complemented with Chebyshev's inequality. In both cases we estimate the value of α corresponding to each experimental subject by using a least squares method, i.e. by minimizing the sum of squared errors between predictions and observations. That is, we consider all the choices made across rounds and periods by each subject (90 in the

[28] There is a one-to-one correspondence between the two. For the 0.7 condition, $p_w = 0.7\ P(Urn1) + 0.3\ P(Urn2) = 0.7\ P(Urn1) + 0.3\ [1 - P(Urn1)]$. A similar expression holds for the 0.6 condition.

[29] For the 0.7 condition, $0.7 \times 0.5 + 0.3 \times 0.5 = 0.5$, and $0.6 \times 0.5 + 0.4 \times 0.5 = 0.5$ for the 0.6 condition.

[30] Naturally, once the Central Limit Theorem holds this is not a shortcut.

[31] On Chebyshev's inequality, see for example Davidson and MacKinnon (1993).

[32] For example, suppose a test of size $\alpha = 0.25$ is required. The above inequality says that the chance of getting an observation more than 2 standard deviations away from the mean is less than 25%. Therefore, if such an observation is observed, a rare event has occurred and the belief can be changed with a chance of making a mistake (probability of a type I error) no greater than 25%.

full sample) and we find the *subject-specific* value of α that minimizes the sum of squared differences between IE (IE_C or IE_N) and such choices.[33]

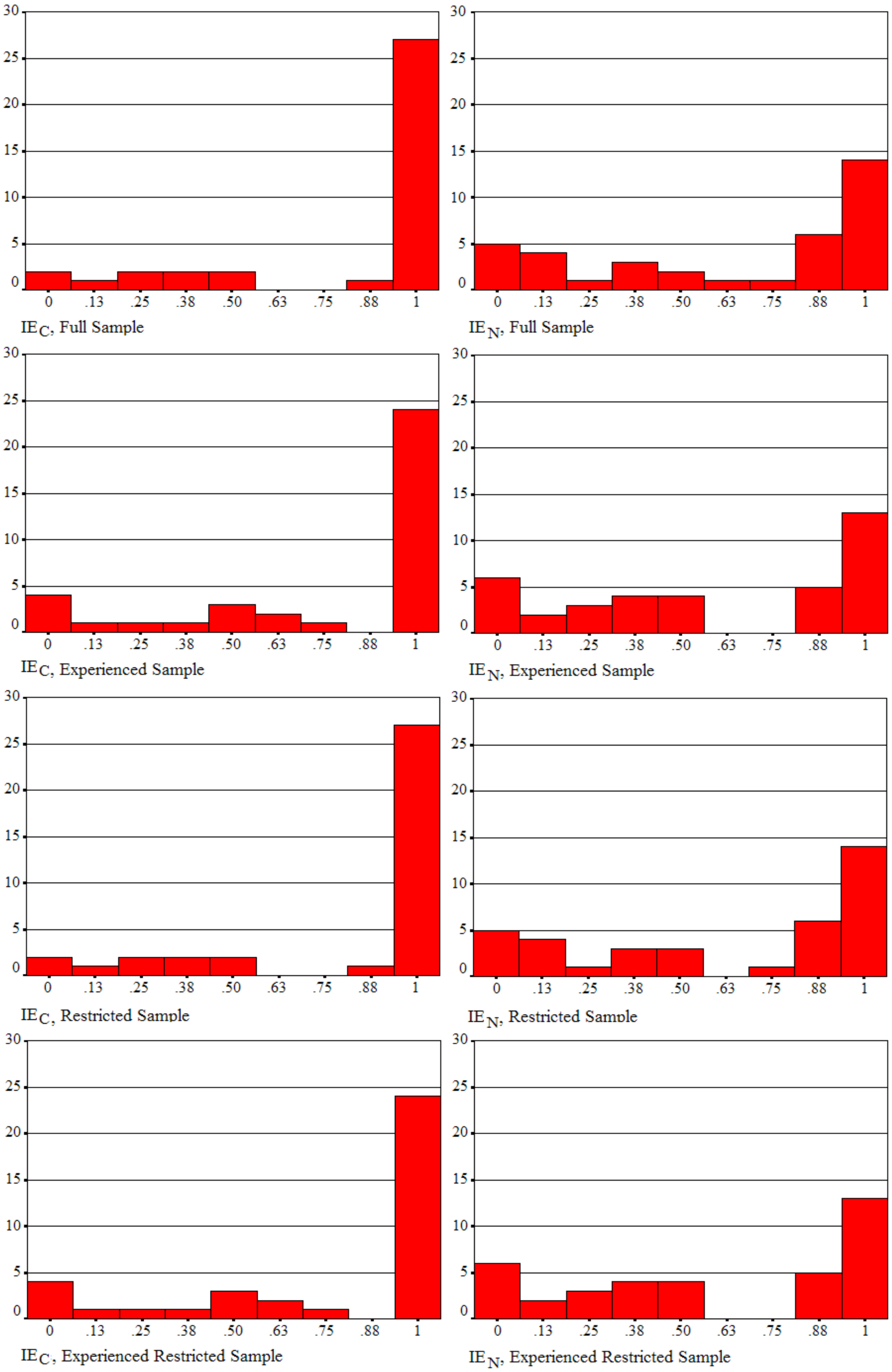
Experimental Hypothesis. We can compute the expectations profile for RE, IE_N, and IE_C agents in relation to each session, using the sequence of observed ball draws and the procedures described so far.
HYPOTHESIS 1. IE_N and IE_C have lower mean square errors from empirically observed choices than RE.
HYPOTHESIS 2. IE_N and IE_C are better able to predict belief conservatism than RE.
Appendix E presents two additional hypotheses designed to test the robustness of IE. In testing Hypothesis 1 and 2 we used not only the 'full' sample from all six periods but also an 'experienced' sample which removes the observations from periods 1 and 2, thus allowing subjects to get some practice and experience about the nature of the task. We also considered a 'restricted' sample of observations where periods in which subjects altered their choices (notwithstanding our instructions to the contrary), and periods where some misperceptions occurred in the recording of the colour of the balls, were removed.[34] Overall, we employed four samples: the full sample, the experienced sample, the restricted sample, and the experienced restricted sample.

5.4 *Experimental Results*

Estimation of α *values*. Figure 3 provides histograms for the distribution of α for both IE_N and IE_C, in the various samples.

[33] Or, equivalently, the mean sum of square error computed by observation; see Appendix D (section D.1) for details.
[34] A total of nine periods were removed in this way, six from choice alteration and three from apparent misperception. Five of the nine periods removed were in periods 1 or 2.

FIGURE 3. HISTOGRAMS OF α VALUES FOR IE



In the experienced (full) sample mean α values were 0.585 (0.635) for IE_N and 0.767 (0.813) for IE_C; mean values in the corresponding restricted samples were virtually identical (see Table 3). Table 4 shows the percentage of subjects displaying $\alpha < 1$ and $\alpha < 0.9$ in the various samples.

TABLE 3 - MEAN α VALUES BY CONDITION

	Full Sample		Experienced Sample	
	Condition		Condition	
	0.6	0.7	0.6	0.7
IE_C	0.819	0.808	0.736	0.796
IE_N	0.703	0.571	0.597	0.573
	Restricted Sample		Exp. Restricted Sample	
	Condition		Condition	
	0.6	0.7	0.6	0.7
IE_C	0.82	0.808	0.736	0.796
IE_N	0.696	0.571	0.597	0.573

Notes: The table contains the mean of the α values estimated for each subject and sample in relation to IE_C and IE_N, by experimental condition. The hypothesis that the means are the same between conditions (see Appendix E) cannot be rejected in Mann-Whitney tests.

TABLE 4 - PERCENTAGE OF SUBJECTS WITH $\alpha < 1$ OR 0.9

		Sample			
		Full	Exp	Restr	Exp+Restr
$\alpha < 1$	IE_C	0.297	0.378	0.297	0.378
	IE_N	0.757	0.784	0.757	0.784
$\alpha < 0.9$	IE_C	0.27	0.351	0.27	0.351
	IE_N	0.568	0.649	0.568	0.649

Notes: The table displays the percentages of subjects, out of $n = 37$, for which $\alpha < 1$ or $\alpha < 0.9$. Full: full sample; Exp: experienced sample; Restr: restricted sample; Exp+Restr: experienced restricted sample. Percentages are computed out of $n = 37$.

A non-negligible fraction of agents had $\alpha < 1$ in both cases: for example, in the experienced sample, 14 out of 37 subjects (0.378) seems to have employed $\alpha < 1$ for IE_C, a number rising to 29 out of 37 (0.757) for IE_N. It is interesting to look also at the $\alpha < 0.9$ fraction of agents, as this may remove α estimation cases which are virtually indistinguishable, in terms of goodness of fit and predictions, from $\alpha = 1$. For IE_C, all but one of the 11-14 subjects for whom $\alpha < 1$ also have $\alpha < 0.9$. For IE_N, over half of the subjects have $\alpha < 0.9$.

The clear differences in the distributions of αs between IE_N and IE_C may suggest that IE_N and IE_C may bear little relation to one another. However, while there were differences in the distributions of the αs, Pearson r (IE_N, IE_C) is equal to 0.887, 0.843, 0.890 and 0.845 in the full, experienced, restricted, and experienced restricted samples respectively[35] ($P < 0.001$). Although IE_N values tend to be lower than IE_C values, IE_N and IE_C predictions tend to follow each other closely.

Appendix E provides additional results about α estimation. We show that α values are robust between conditions and, if estimated on a period-by-period basis, they do not tend to converge to 1, i.e. to the RE benchmark. Another way of looking at the data is to constrain α to take only one possible value below 1, so as to classify subjects in one of just two categories, RE holders and IE (with $\alpha < 1$) holders. Table 5 illustrates the results of this exercise.

TABLE 5 - CONSTRAINED α ESTIMATION

Sample		α	%IE
Full	IE_C	0.27	16.216
	IE_N	0.05	16.216
Experienced	IE_C	0.01-0.09	16.22
	IE_N	0.04	21.621
Restricted	IE_C	0.27	16.216
	IE_N	0.05	16.216
Experienced Restricted	IE_C	0.01-0.09	16.22
	IE_N	0.04	21.621

Notes: The table displays the estimated value of α if α is allowed to take only one value less than 1, and the percentage of subjects for which $\alpha < 1$ correspondingly applies. In two cases α is expressed as a range since any value of α within this range has an equally good fit.

Even under the constraint of only one value $\alpha < 1$, between 16 and 22% are estimated to have IE, and α values are around 0.05 for IE_N and for two out of four samples for IE_C.[36]

Hypothesis 1. We test Hypothesis 1 by computing the mean square error (MSE) values between choices and predictions according to each algorithm. One problem in doing so is the very likely non-independence of observations made by each subject. To address this problem, we use MSE values computed by observation, but only as the basis for a nonparametric sign test, with the number of subjects $n = 37$ as the degrees of freedom.[37] The results of the sign tests[38] are illustrated in Table 6.

(*Table 6 appears at the end of the document.*)

IE_N outperforms IE_C: for example, in the experienced sample, out of 37 subjects IE_N performed worse than IE_C twice, tied for twenty-two subjects, and performed better 14 times

---

[35] Spearman ρ (IE_N, IE_C) is equal to 0.822, 0.774, 0.831 and 0.775 in the full, experienced, restricted, and experienced restricted samples respectively ($P < 0.001$).

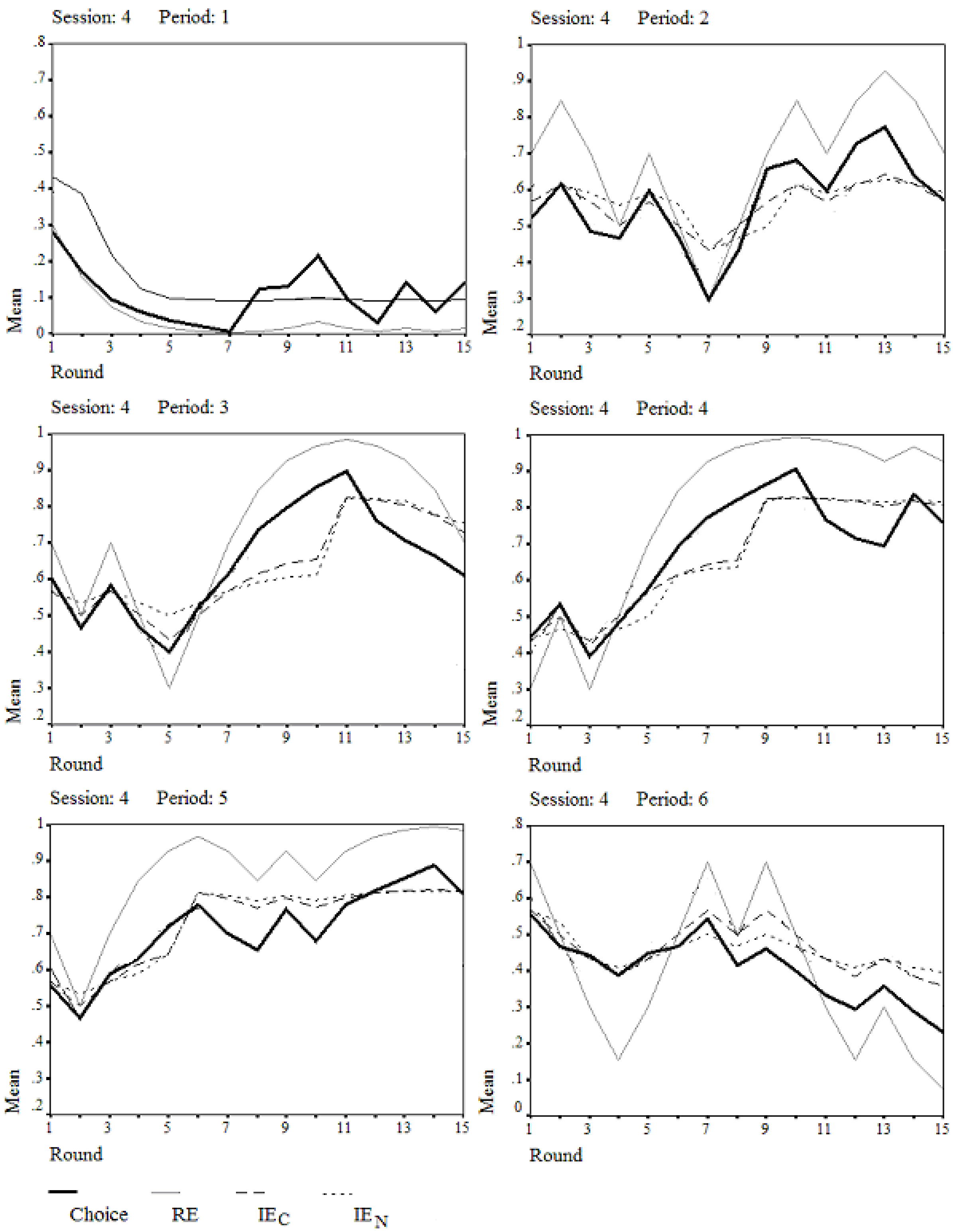
[36] In the rest of this section we use the MSE-minimizing α values, rather than these constrained values.

[37] We perform robustness checks by running sign tests also on MSE values computed by period and on MSE values computed by subject. See Appendix D for details.

[38] We cannot use Wilcoxon tests because Wilcoxon tests rely on the assumption that the sample is drawn from a symmetric distribution, assumption which is clearly violated in this case. For results of (also unsatisfactory) F tests, see Menzies and Zizzo (2004).

(P = 0.007, two-tailed). An electronically available appendix[39] contains the mean choices and predictions according to each model of expectation formation by session and period; Figure 4 exemplifies the kind of aggregate dynamics observed by reproducing the graphs from session 4 (a 0.7 condition session).

FIGURE 4. FULL SAMPLE MEAN CHOICES AND PREDICTIONS: EXAMPLES



[39] The appendix is at http://www.uea.ac.uk/~ec601/MZ/MenziesZizzoWebAppendix.pdf.

Notes: the lines for IE_C and IE_N overlap in period 1.

In period 1 RE performs better in the first seven rounds, but IE does better on average afterwards. In periods 2 and 6, IE may be doing a better job in capturing the lower variability of choice relative to RE. In periods 3, 4 and 5 IE clearly does a better job at tracking mean choices than RE. Relative to RE, there appears to be a lower mean sensitivity of IE (with $\alpha < 1$) predictions to new information (though exceptions exist).

Hypothesis 2. RE predicts that agents should revise their guesses every round in response to new information. This is not the case: Table 7 reports that 35% of the times subjects chose not to switch guesses.

TABLE 7 - NO SWITCH ANALYSIS

Full Sample		Restricted Sample	
Observed	0.354	Observed	0.356
IE_C	0.307	IE_C	0.307
IE_N	0.365	IE_N	0.367
$\rho(IE_C)$	0.027	$\rho(IE_C)$	0.096
$\rho(IE_N)$	0.364*	$\rho(IE_N)$	0.239°
Experienced Sample		**Experienced Restricted Sample**	
Observed	0.354	Observed	0.354
IE_C	0.363	IE_C	0.346
IE_N	0.424	IE_N	0.425
$\rho(IE_C)$	0.027	$\rho(IE_C)$	-0.19
$\rho(IE_N)$	0.364*	$\rho(IE_N)$	0.390**

Notes: Observed, IE_C and IE_N are the fractions of not switching choices (i.e., not different from previous round in same period) respectively observed or predicted by IE_C and IE_N. $\rho(IE_C)$ and $\rho(IE_N)$ are the respective Spearman correlation coefficients of the percentage of no switches predicted for each subject by IE_C and IE_N with the observed percentage of no switches.°: significant at the 0.1 level; *: significant at the 0.05 level; **: significant at the 0.01 level.

This is very close to the mean predictions of IE_N and IE_C (with MSE-minimizing α values), which depending on the sample chosen range between 30% and 42% (the difference is never significant in nonparametric Wilcoxon tests[40]). Spearman correlation coefficients between the mean amount of 'no switches' observed and predicted by IE_N and IE_C show, however, that, while IE_N has predictive power, IE_C has not.

5. Discussion and Conclusions

This paper has presented a new model of belief formation. The basic idea of IE is that beliefs are maintained or revised using a Neyman-Pearson hypothesis test. They are rejected in favour of RE only when the rejection region, determined by the test size α, is reached. The

[40] Sign tests yield the same answer.

nesting of RE within IE means that the estimated α becomes a metric for the 'rationality' – the closeness to RE beliefs - of agents.

This fast and frugal heuristic is consistent with the scientific practices of most scientists, and with a view of decision-making characterized by information-gathering and information-processing costs. It is also consistent with the view that agents do not pay attention all the time to new flows of information, and only when a threshold is reached do they pay attention and switch beliefs. Our experimental setup was a fairly hostile one for inferential expectations since the only task subjects had to do was to choose probabilities, and hence attention is unlikely to have been a serious issue: this may under-estimate the use of inferential expectations. Yet, our experimental evidence suggested that between one and two thirds of agents exhibited α's less than 0.9, and that there is significant evidence of belief conservatism. When we constrained α to take only one possible value other than 1, we found that for IE complemented by the normal approximation method this value was in the region of a 0.05 significance level for around one-fifth of the subjects.

We presented a variant of the Dornbusch (1976) model, and showed that simply replacing RE with IE can explain the empirical failures of regressions on uncovered interest parity and on the term structure of interest rates. Naturally, these results may be contingent upon the specification of the null and alternative hypotheses, the assumption of a common α across all agents and the macro-model within which the exchange rate and long interest rate are imbedded. Nevertheless, we suggest that our approach may be valuable, since it attributes the extent of parameter bias to deep model parameters, rather than risk premia.

We intend to develop IE further, both theoretically and empirically. In future work we hope to place IE into full scale macroeconomic models, to see if equations *(1)* and *(2)* continue to display parameter bias. Inferential expectations could also be relevant for more sophisticated hypothesis testing, for example to detect the presence of autocorrelation, as in Rotheli (1998). Finally, we have confined our attention to a single change in belief. If continual inference on the long-run exchange rate meant repeated changes in beliefs, the variance of the current exchange rate would be dominated by 'shocks' (really belief changes) to the long-run rate (Campbell and Clarida, 1987). The goal of this paper has simply been to define IE, present some evidence in favour of it, and demonstrate its potential fruitfulness as a modelling device.

APPENDIX A. EXCHANGE RATE MODEL

A.1 DERIVATION OF EQUATIONS (7)-(9)

With the normalizations in the text, the RE system becomes:

$$m - p_t = -B_i i_t \qquad (3')$$
$$p_{t+1} - p_t = -B_p(p_t - \Delta m). \qquad (4')$$
$$s_{t+1} - s_t = i_t. \qquad (5')$$

The steady-state level of prices in *(4')* is Δm from the properties of the system in the steady state. As $\Delta s=0=i$ (from *(5')*), $p=m=\Delta m$ (from *(3')*). Furthermore, from purchasing power parity (which holds in the steady state) $s=p=\Delta m$. We now solve the system.

from (4)
$$p_{t+1} - \Delta m = (1 - B_p)(p_t - \Delta m)$$
$$p_t - \Delta m = (p_{0+} - \Delta m)(1 - B_p)^t$$
$$= -\Delta m(1 - B_p)^t \qquad \textit{which is } (7)$$

p_{0+} is zero from the Dornbusch assumption of sticky prices ($p_{0+}=p_{0-}$). The eigenvalue of the system is clearly ($1-B_p$). To find i, we substitute $m=\Delta m$ and *(7)* into money demand.

$$\Delta m - p_t = -B_i i_t$$
$$i_t = \frac{p_t - \Delta m}{B_i} = \frac{-\Delta m(1 - B_p)^t}{B_i} \qquad \textit{which is } (8)$$

Clearly, $i_{t+1}=(1-B_p)i_t$. To obtain s, *(5')* is iterated forward to infinity (when $s=\Delta m$) and the infinite GP, with ratio $(1-B_p)$, is summed.

$$s_t = -(i_t + i_{t+1} + i_{t+2} \ldots) + s_\infty$$
$$= -(i_t + (1 - B_p)i_t + (1 - B_p)^2 i_t + \ldots) + \Delta m$$
$$= \frac{-i_t}{B_p} + \Delta m \qquad \textit{which is } (9)$$

A.2 COHERENCE OF BELIEFS ON THE CENTRAL BANK POLICY

This part of the appendix demonstrates that the central bank policy of driving all future log-exchange rates to zero (i.e. $s = 0$) is achievable, even though the price level has already changed (for $t > 0$).

First, by solving *(5')* forwards and assuming $s_\infty=0$, it is clear that setting all future i to zero will force all future s to zero (solving back from the long run). Setting $i=0$ in the money demand equation *(3')* requires $m=p$. How is p determined? Again assuming $s_\infty=0$ purchasing power parity implies $p_\infty=0$. Thus, if steady-state p is zero, equation *(4')* becomes $p_{t+1}=(1-B_p)p_t$ or $p_z=p_t(1-B_p)^{z-t}$, where z ($> t$) refers to all future periods from t.

Thus, IE agents who understand the model and who think that the central bank's desired log-exchange rate is zero (including, crucially, s_{∞}=0) will believe, coherently with these beliefs, that at period t+1 the money stock (or the nominal income target if *(3')* is a Taylor rule) will jump back to $p_{t+1}=p_t(1-B_p)$ so that interest rates will be zero (since m-p will be zero); and that, from t+1 forward, m will equal p and both will converge to zero at rate $(1-B_p)$.

Such a policy implies that the central bank is permitting only a temporary change to the price level (p_{∞}=0 as noted in the text), and that all future interest differentials are zero (as noted in the statement of H_0).

A.3 Interest Rate Forecasting

It was asserted in the text that, under IE, a regression of the expectation of i_{t+1} (at time t) on current i_t (without a constant term) can yield a high autocorrelation parameter for low values of B_p and high values of α. We present here the proof of this statement (related computer simulations can be found in Appendix B).

Proof: The OLS coefficient numerator is divided into two parts, with one part vanishing.

$$\hat{\theta} = \frac{\sum_{t=0}^{t^*-1} {}^{H0}i_{t+1}i_t + \sum_{t=t^*1}^{n} i_{t+1}i_t}{\sum_{t=0}^{n} i_t^{\,2}} = \frac{\sum_{t=t^*}^{n} i_{t+1}i_t}{\sum_{t=0}^{n} i_t^{\,2}} = \frac{(1-B_p)\sum_{t=t^*}^{n} i_t^{\,2}}{\sum_{t=0}^{n} i_t^{\,2}}$$

Clearly, a high α (implying a low t^*) will make this coefficient approximately equal to *(1-B_p)*. That being so, a low B_p implies a high (close to unity) autocorrelation. If we are prepared to assume a large sample, and define t* according to the rejection region *(10)* for p_t:

$$\mathrm{p}_{\mathrm{t}^*} \leq (1-\alpha)\,\Delta m \qquad \Rightarrow \qquad (1-B_p)^{t^*} \leq \alpha$$

we can further simplify the OLS parameter using *(8)*.

$$\hat{\theta} = \frac{(1-B_p)\sum_{t=t^*}^{\infty} i_t^{\,2}}{\sum_{t=0}^{\infty} i_t^{\,2}} \leq (1-B_p)\;\alpha^2$$

With finely graduated time intervals, the inequalities will be approximate equalities, and a low B_p and a high α will result in a high OLS coefficient. *Q.E.D.*

Appendix B. Exchange Rate Model Simulations

810 versions of the model were generated with all the combinations of the following parameter values for β_p, β_i and α: β_p = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]; β_i = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]; α = [0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]. We then ran six sets of regressions for each of this version. Two sets are the uncovered interest rate

regressions (equation *(1)* in the main text), with and without the constant term. The third and fourth sets correspond to the term structure regressions (equation *(2)* in the main text), with and without the constant term. The last two sets are interest rate forecasting regressions (under IE), again with and without the constant term:

$$\Delta s_{t+1} = \beta_1 + \beta_2 \ (i_t - i_t^*) + u_t$$
$$\Delta s_{t+1} = \beta_2 \ (i_t - i_t^*) + u_t$$
$$\frac{\Delta i_{t+1}}{2} = \gamma_1 + \gamma_2 \ (r_t - i_t) + v_t$$
$$\frac{\Delta i_{t+1}}{2} = \gamma_2 \ (r_t - i_t) + v_t$$
$$E(i_{t+1}) = \theta_1 + \theta_2 \ i_t + v_t$$
$$E(i_{t+1}) = \theta_2 \ i_t + v_t$$

UIP regressions. In the main text, Table 1 considered the UIP regressions with constant, with mean slope coefficient values classified by α and β_p. Table 8 is also on the UIP regressions. We find that the UIP slope coefficient is robustly lower than 1 both in regressions with and without slope. It is lower for lower β_i and α values and the higher the β_i. Overall, the downwards bias shows a considerable degree of robustness, especially for realistic values of β_p (0.1-0.3), equivalent to 75% of the impact on inflation occurring between 1 and 3 years.

TABLE 8 – UNCOVERED INTEREST RATE REGRESSIONS
A. WITHOUT CONSTANT: MEAN SLOPE COEFFICIENT VALUES CLASSIFIED BY α AND β_p

	α									
β_p	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.091	0.079	0.045	-0.011	-0.093	-0.181	-0.313	-0.402	-0.511	-0.808
	(0.074)	(0.111)	(0.175)	(0.246)	(0.322)	(0.387)	(0.467)	(0.513)	(0.564)	(0.681)
0.2	0.188	0.173	0.134	0.072	0.019	-0.06	-0.177	-0.354	-0.62	-0.62
	(0.074)	(0.093)	(0.13)	(0.172)	(0.201)	(0.237)	(0.281)	(0.333)	(0.394)	(0.394)
0.3	0.284	0.263	0.211	0.155	0.057	-0.119	-0.119	-0.445	-0.445	-0.445
	(0.078)	(0.091)	(0.117)	(0.139)	(0.17)	(0.21)	(0.21)	(0.259)	(0.259)	(0.259)
0.4	0.373	0.352	0.314	0.238	0.078	0.078	-0.28	-0.28	-0.28	-0.28
	(0.083)	(0.09)	(0.101)	(0.119)	(0.146)	(0.146)	(0.175)	(0.175)	(0.175)	(0.175)
0.5	0.476	0.449	0.391	0.25	0.25	-0.125	-0.125	-0.125	-0.125	-0.125
	(0.084)	(0.089)	(0.098)	(0.112)	(0.112)	(0.119)	(0.119)	(0.119)	(0.119)	(0.119)
0.6	0.572	0.529	0.406	0.406	0.02	0.02	0.02	0.02	0.02	0.02
	(0.086)	(0.09)	(0.096)	(0.096)	(0.078)	(0.078)	(0.078)	(0.078)	(0.078)	(0.078)
0.7	0.658	0.555	0.555	0.155	0.155	0.155	0.155	0.155	0.155	0.155
	(0.089)	(0.09)	(0.09)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)
0.8	0.702	0.702	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
	(0.089)	(0.089)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
0.9	0.85	0.395	0.395	0.395	0.395	0.395	0.395	0.395	0.395	0.395
	(0.091)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)

Notes: The table reports mean slope coefficients (and mean standard errors in parenthesis) across simulations with different values of β_i.

B. WITH CONSTANT: MEAN SLOPE COEFFICIENT VALUES CLASSIFIED BY α AND β_i

	α									
β_i	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.502	0.48	0.442	0.388	0.319	0.242	0.163	0.081	0.018	-0.055
	(0.029)	(0.041)	(0.063)	(0.089)	(0.116)	(0.141)	(0.165)	(0.186)	(0.202)	(0.221)
0.2	0.501	0.464	0.411	0.344	0.265	0.181	0.097	0.012	-0.053	-0.127
	(0.045)	(0.055)	(0.075)	(0.1)	(0.125)	(0.149)	(0.173)	(0.193)	(0.21)	(0.228)
0.3	0.5	0.447	0.379	0.299	0.21	0.119	0.03	-0.058	-0.124	-0.199
	(0.061)	(0.068)	(0.088)	(0.11)	(0.134)	(0.157)	(0.181)	(0.201)	(0.217)	(0.235)
0.4	0.5	0.431	0.348	0.255	0.156	0.058	-0.036	-0.127	-0.195	-0.271
	(0.076)	(0.082)	(0.1)	(0.121)	(0.144)	(0.166)	(0.188)	(0.209)	(0.225)	(0.243)
0.5	0.499	0.414	0.316	0.21	0.101	-0.004	-0.102	-0.197	-0.267	-0.343
	(0.092)	(0.096)	(0.112)	(0.132)	(0.153)	(0.174)	(0.196)	(0.216)	(0.232)	(0.25)
0.6	0.498	0.398	0.285	0.166	0.047	-0.066	-0.168	-0.266	-0.338	-0.415
	(0.108)	(0.11)	(0.124)	(0.142)	(0.162)	(0.183)	(0.204)	(0.224)	(0.24)	(0.258)
0.7	0.497	0.382	0.253	0.122	-0.008	-0.127	-0.235	-0.336	-0.409	-0.487
	(0.124)	(0.124)	(0.136)	(0.153)	(0.172)	(0.191)	(0.212)	(0.232)	(0.247)	(0.265)
0.8	0.496	0.365	0.222	0.077	-0.062	-0.189	-0.301	-0.406	-0.48	-0.559
	(0.14)	(0.138)	(0.149)	(0.164)	(0.181)	(0.2)	(0.22)	(0.239)	(0.254)	(0.272)
0.9	0.495	0.349	0.19	0.033	-0.117	-0.251	-0.367	-0.475	-0.551	-0.631
	(0.156)	(0.152)	(0.161)	(0.174)	(0.191)	(0.208)	(0.228)	(0.247)	(0.262)	(0.28)

Notes: The table reports mean slope coefficients (and mean standard errors in parenthesis) across simulations with different values of β_p.

C. WITHOUT CONSTANT: MEAN SLOPE COEFFICIENT VALUES CLASSIFIED BY α AND β_i

	α									
β_i	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.492	0.473	0.442	0.4	0.349	0.293	0.24	0.184	0.146	0.115
	(0.025)	(0.033)	(0.051)	(0.071)	(0.091)	(0.111)	(0.13)	(0.147)	(0.16)	(0.173)
0.2	0.486	0.452	0.407	0.354	0.294	0.232	0.175	0.117	0.078	0.046
	(0.039)	(0.046)	(0.061)	(0.08)	(0.099)	(0.118)	(0.137)	(0.154)	(0.166)	(0.179)
0.3	0.479	0.431	0.373	0.308	0.239	0.171	0.111	0.05	0.01	-0.022
	(0.054)	(0.058)	(0.072)	(0.089)	(0.108)	(0.126)	(0.143)	(0.16)	(0.173)	(0.186)
0.4	0.473	0.41	0.338	0.262	0.184	0.11	0.046	-0.017	-0.058	-0.09
	(0.068)	(0.071)	(0.083)	(0.098)	(0.116)	(0.133)	(0.15)	(0.167)	(0.18)	(0.193)
0.5	0.466	0.389	0.303	0.216	0.129	0.049	-0.018	-0.084	-0.126	-0.159
	(0.083)	(0.084)	(0.094)	(0.108)	(0.124)	(0.14)	(0.157)	(0.174)	(0.186)	(0.199)
0.6	0.46	0.368	0.269	0.169	0.074	-0.012	-0.083	-0.151	-0.194	-0.227
	(0.098)	(0.097)	(0.105)	(0.117)	(0.132)	(0.148)	(0.164)	(0.18)	(0.193)	(0.206)
0.7	0.453	0.347	0.234	0.123	0.019	-0.072	-0.147	-0.218	-0.261	-0.295
	(0.112)	(0.109)	(0.116)	(0.127)	(0.14)	(0.155)	(0.171)	(0.187)	(0.199)	(0.212)
0.8	0.447	0.326	0.199	0.077	-0.035	-0.133	-0.212	-0.285	-0.329	-0.364
	(0.127)	(0.122)	(0.127)	(0.136)	(0.148)	(0.162)	(0.178)	(0.194)	(0.206)	(0.219)
0.9	0.44	0.305	0.165	0.031	-0.09	-0.194	-0.276	-0.352	-0.397	-0.432
	(0.142)	(0.135)	(0.138)	(0.145)	(0.157)	(0.17)	(0.185)	(0.2)	(0.212)	(0.225)

Notes: The table reports mean slope coefficients (and mean standard errors in parenthesis) across simulations with different values of β_p.

Term structure regressions. In the main text, Table 2 considered the term structure regressions with constant, with mean slope coefficient values classified by α and β_p. Table 9 is also on the term structure regressions. As predicted by equation (16), the slope coefficient is lower with lower β_p values, leading to small mean coefficients for $\beta_p \leq 0.3$. Slope coefficients are very robust to the presence of a constant and to different α values, and entirely invariant to changes in β_i.

TABLE 9 – TERM STRUCTURE REGRESSIONS

A. WITHOUT CONSTANT: MEAN SLOPE COEFFICIENT VALUES CLASSIFIED BY α AND β_p

	α									
β_p	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.100	0.100	0.100	0.101	0.102	0.103	0.105	0.107	0.110	0.137
	(0)	(0.001)	(0.002)	(0.003)	(0.005)	(0.006)	(0.008)	(0.009)	(0.012)	(0.022)
0.2	0.200	0.200	0.201	0.202	0.204	0.206	0.211	0.222	0.253	0.253
	(0.001)	(0.002)	(0.003)	(0.005)	(0.007)	(0.009)	(0.011)	(0.016)	(0.024)	(0.024)
0.3	0.300	0.300	0.302	0.304	0.308	0.319	0.319	0.356	0.356	0.356
	(0.001)	(0.002)	(0.004)	(0.006)	(0.009)	(0.014)	(0.014)	(0.023)	(0.023)	(0.023)
0.4	0.400	0.401	0.402	0.405	0.414	0.414	0.450	0.450	0.450	0.450
	(0.001)	(0.002)	(0.004)	(0.006)	(0.011)	(0.011)	(0.02)	(0.02)	(0.02)	(0.02)
0.5	0.500	0.500	0.502	0.508	0.508	0.538	0.538	0.538	0.538	0.538
	(0.001)	(0.002)	(0.004)	(0.008)	(0.008)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
0.6	0.600	0.601	0.604	0.604	0.626	0.626	0.626	0.626	0.626	0.626
	(0.001)	(0.002)	(0.005)	(0.005)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
0.7	0.700	0.701	0.701	0.714	0.714	0.714	0.714	0.714	0.714	0.714
	(0.001)	(0.002)	(0.002)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
0.8	0.800	0.800	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805
	(0.001)	(0.001)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
0.9	0.900	0.901	0.901	0.901	0.901	0.901	0.901	0.901	0.901	0.901
	(0)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Notes: The table reports mean slope coefficients (and mean standard errors in parenthesis) across simulations with different values of β_i.

B. MEAN SLOPE COEFFICIENT VALUES CLASSIFIED BY α AND β_i

	α									
Constant	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Yes	0.499	0.498	0.498	0.499	0.501	0.505	0.508	0.512	0.515	0.516
	(0.001)	(0.002)	(0.003)	(0.005)	(0.007)	(0.009)	(0.01)	(0.012)	(0.013)	(0.014)
No	0.500	0.501	0.502	0.505	0.509	0.514	0.519	0.524	0.528	0.531
	(0.001)	(0.002)	(0.003)	(0.005)	(0.007)	(0.009)	(0.01)	(0.012)	(0.013)	(0.014)

Notes: The table reports mean slope coefficients (and mean standard errors in parenthesis) across simulations with different values of β_p and β_i. Coefficients and standard errors are invariant to different values of β_i.

Interest forecasting regressions. Tables 10 through 12 consider interest rate forecasting equations. They show that, for $\beta_p \leq 0.3$ and high values of α, the mean slope coefficient can be positive and larger than 0.

TABLE 10 – INTEREST RATE FORECASTING REGRESSIONS WITH CONSTANT: MEAN SLOPE COEFFICIENT VALUES CLASSIFIED BY α AND β_p

	α									
β_p	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	-0.009	-0.057	-0.112	-0.137	-0.119	-0.067	0.045	0.135	0.256	0.627
	(0.007)	(0.017)	(0.029)	(0.045)	(0.065)	(0.083)	(0.101)	(0.11)	(0.115)	(0.097)
0.2	-0.012	-0.019	-0.023	-0.008	0.014	0.055	0.128	0.253	0.46	0.46
	(0.006)	(0.012)	(0.025)	(0.039)	(0.049)	(0.06)	(0.072)	(0.083)	(0.083)	(0.083)
0.3	-0.005	-0.007	-0.002	0.012	0.047	0.129	0.129	0.31	0.31	0.31
	(0.005)	(0.011)	(0.022)	(0.031)	(0.044)	(0.059)	(0.059)	(0.07)	(0.07)	(0.07)
0.4	-0.003	-0.003	0	0.013	0.057	0.057	0.195	0.195	0.195	0.195
	(0.005)	(0.009)	(0.014)	(0.024)	(0.039)	(0.039)	(0.056)	(0.056)	(0.056)	(0.056)
0.5	-0.001	-0.001	0.002	0.021	0.021	0.112	0.112	0.112	0.112	0.112
	(0.003)	(0.006)	(0.012)	(0.023)	(0.023)	(0.042)	(0.042)	(0.042)	(0.042)	(0.042)
0.6	-0.001	0	0.006	0.006	0.056	0.056	0.056	0.056	0.056	0.056
	(0.002)	(0.005)	(0.012)	(0.012)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
0.7	0	0.001	0.001	0.023	0.023	0.023	0.023	0.023	0.023	0.023
	(0.002)	(0.005)	(0.005)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
0.8	0	0	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
	(0.001)	(0.001)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
0.9	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(0)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

Notes: The table reports mean slope coefficients (and mean standard errors in parenthesis) across simulations with different values of β_i.

TABLE 11 – INTEREST RATE FORECASTING REGRESSIONS WITHOUT CONSTANT: MEAN SLOPE COEFFICIENT VALUES CLASSIFIED BY α AND β_p

	α									
β_p	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.001	0.007	0.03	0.071	0.134	0.205	0.313	0.387	0.478	0.729
	(0.005)	(0.015)	(0.029)	(0.044)	(0.058)	(0.069)	(0.078)	(0.081)	(0.082)	(0.064)
0.2	0.002	0.006	0.023	0.055	0.086	0.134	0.21	0.328	0.512	0.512
	(0.006)	(0.012)	(0.024)	(0.037)	(0.045)	(0.055)	(0.064)	(0.072)	(0.07)	(0.07)
0.3	0.001	0.005	0.02	0.04	0.082	0.168	0.168	0.343	0.343	0.343
	(0.005)	(0.01)	(0.021)	(0.03)	(0.041)	(0.055)	(0.055)	(0.064)	(0.064)	(0.064)
0.4	0.001	0.004	0.01	0.028	0.078	0.078	0.216	0.216	0.216	0.216
	(0.005)	(0.008)	(0.014)	(0.023)	(0.037)	(0.037)	(0.053)	(0.053)	(0.053)	(0.053)
0.5	0	0.002	0.008	0.031	0.031	0.125	0.125	0.125	0.125	0.125
	(0.003)	(0.006)	(0.011)	(0.022)	(0.022)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
0.6	0	0.002	0.01	0.01	0.064	0.064	0.064	0.064	0.064	0.064
	(0.002)	(0.005)	(0.012)	(0.012)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
0.7	0	0.002	0.002	0.027	0.027	0.027	0.027	0.027	0.027	0.027
	(0.001)	(0.005)	(0.005)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
0.8	0	0	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
	(0.001)	(0.001)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
0.9	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(0)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

Notes: The table reports mean slope coefficients (and mean standard errors in parenthesis) across simulations with different values of β_i.

TABLE 12 – INTEREST RATE FORECASTING REGRESSIONS WITH AND WITHOUT CONSTANT: MEAN SLOPE COEFFICIENT VALUES CLASSIFIED BY α AND β_i

	α									
Constant	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Yes	-0.003	-0.01	-0.014	-0.007	0.012	0.041	0.077	0.121	0.158	0.199
	(0.003)	(0.007)	(0.014)	(0.022)	(0.03)	(0.037)	(0.043)	(0.046)	(0.047)	(0.045)
No	0.001	0.003	0.012	0.03	0.057	0.09	0.126	0.166	0.197	0.225
	(0.003)	(0.007)	(0.014)	(0.021)	(0.028)	(0.034)	(0.038)	(0.04)	(0.04)	(0.038)

Notes: The table reports mean slope coefficients (and mean standard errors in parenthesis) across simulations with different values of β_p and β_i. Coefficients and standard errors are invariant to different values of β_i.

APPENDIX C. EXPERIMENTAL INSTRUCTIONS

Instructions for 0.7 Condition

Welcome to the experiment!

The experiment is divided into two parts, *Stage 1* and *Stage 2*. Your final winnings will be equal to the Stage 1 Payment, the Stage 2 Payment and a participation fee of 8 dollars. (All winnings will be rounded to the nearest 5 cents).

You are playing Stage 1 first. You can see two identical urns on the table, and a set of white and orange balls in a basket; you can also see a screen. The experimenter will shortly do the following:
(a) show you that the urns are empty;
(b) take **seven white balls** and **three orange balls**, and put them in one of the two urns; let us label this urn **Urn 1**;
(c) take **three white balls** and **seven orange balls**, and put them in the other urn; let us label this urn **Urn 2**;
(d) hide both urns behind the screen.

There are six *periods* in Stage 1. You have received an answer booklet with a sheet for each period.

At the start of each period, the experimenter announces the period number and writes it on the board. Then one of the two urns will be randomly chosen, by the flip of a coin, independently of what urns were chosen in previous periods. You will not be able to see whether this *chosen urn* is Urn 1 or Urn 2, but you will be asked to guess how likely you think it is that the chosen urn is Urn 1.

There are *sixteen draws* in each period. At the start of each draw the experimenter announces the draw number and writes it on the board. *In Draw* 0, which happens at the start of the period, your best probability guess that the chosen urn is Urn 1 would have to be 50%: this is because *at the start* of each period the chosen urn is picked randomly afresh. This Draw 0 probability guess has been printed into the answer booklet for you.

For Draws 1 through Draw 15 inclusive:

1. *first*, the experimenter draws a ball from the chosen urn and announces whether it is white or orange; *please write the ball colour on the answer sheet, in the line corresponding to the correct period and draw*; the experimenter then puts the ball back into the chosen urn;

2. *second*, you have to answer the following question: "how likely is it that the chosen urn is Urn 1? (Remember, Urn 1 is the urn with 7 white and 3 orange balls). Please choose a probability over the range 0% (definitely not) to 100% (definitely certain)"; *please put your guess in the line in the answer booklet corresponding to the correct period and draw.*

At the end of the period the experimenter hides the chosen urn again behind the screen. If you are in periods 1 through 5, you should move on to the answer sheet for the following period. If you are in period 6, please wait until the sheets are collected and the material for Stage 2 is distributed.

Stage 1 Payment. It is important that you try to make your best probability guesses, both because it is important for the value of the experiment, and because your final winnings depend on it. At the end of the experiment the experimenter will randomly choose a winning draw to reward your performance. The experimenter will roll a die to choose the period, and pick randomly from a third urn (with balls numbered between 1 through 15) to choose the winning draw. Your Stage 1 Payment will depend on your choice in the draw corresponding to the number on the ball which has been picked. In relation to this draw, the experimenter will take your choice and compare it with the true chosen urn for that draw. If in the winning draw the chosen urn was Urn 1, then the correct probability of the chosen urn being Urn 1 is 100%; if the chosen urn was Urn 2, then the correct probability of the chosen urn being Urn 1 is 0%. Your Stage 1 Payment will then be equal to

$$25 - 25 \times (\textit{guess} - \textit{correct probability})^2$$

that is, to 25 dollars minus a penalty. The penalty will be equal to the square of the error, that is of difference between the guess and the correct probability, multiplied by 25. The Stage 1 Payment will be higher the more correct your guess is. The enclosed table provides Stage 1 Payment values corresponding to some possible error levels.

Please stay seated throughout the experiment. It is essential, for the scientific value of the experiment, that you (a) do **not** communicate in any way with other participants during the experiment; (b) do **not** change your guesses for previous draws. You are liable to be expelled from the experiment, and forfeit all winnings (including the participation fee), if you do not comply with these simple rules.

This is an individual choice experiment: your choices have no influence on the winnings of other participants, and similarly the choices of other participants have no influence on your winnings. If you have any question, please raise your hand until an experimenter comes close to you, and then ask with a low voice. This may be a good time to ask questions, but feel free to raise your hand to ask questions at any time.

Stage 1 Payment Table

Payment = 25 – 25 × (*guess – correct probability*)^2

Error	Stage 1 Payment	Error	Stage 1 Payment	Error	Stage 1 Payment
0%	25	34%	22.11	68%	13.44
1%	25	35%	21.94	69%	13.1
2%	24.99	36%	21.76	70%	12.75
3%	24.98	37%	21.58	71%	12.4
4%	24.96	38%	21.39	72%	12.04
5%	24.94	39%	21.2	73%	11.68
6%	24.91	40%	21	74%	11.31
7%	24.88	41%	20.8	75%	10.94
8%	24.84	42%	20.59	76%	10.56
9%	24.8	43%	20.38	77%	10.18
10%	24.75	44%	20.16	78%	9.79
11%	24.7	45%	19.94	79%	9.4
12%	24.64	46%	19.71	80%	9
13%	24.58	47%	19.48	81%	8.6
14%	24.51	48%	19.24	82%	8.19
15%	24.44	49%	19	83%	7.78
16%	24.36	50%	18.75	84%	7.36
17%	24.28	51%	18.5	85%	6.94
18%	24.19	52%	18.24	86%	6.51
19%	24.1	53%	17.98	87%	6.08
20%	24	54%	17.71	88%	5.64
21%	23.9	55%	17.44	89%	5.2
22%	23.79	56%	17.16	90%	4.75
23%	23.68	57%	16.88	91%	4.3
24%	23.56	58%	16.59	92%	3.84
25%	23.44	59%	16.3	93%	3.38
26%	23.31	60%	16	94%	2.91
27%	23.18	61%	15.7	95%	2.44
28%	23.04	62%	15.39	96%	1.96
29%	22.9	63%	15.08	97%	1.48
30%	22.75	64%	14.76	98%	0.99
31%	22.6	65%	14.44	99%	0.5
32%	22.44	66%	14.11	100%	0
33%	22.28	67%	13.78		

Answer Booklet: Content of the Sheet for Each Period

When the experimenter draws a ball, write down the colour of the drawn ball in the middle column (if you find it convenient, you can just write W for white and O for orange).

How likely is it that the chosen urn is Urn 1? (Remember, Urn 1 is the urn with 7 white and 3 orange balls). Please choose a probability over the range 0% (definitely not) to 100% (definitely certain). Write down your answer in the Probability Guess column.

Do **not** change probability guesses corresponding to previous draws. If you do, you are liable to to be expelled from the experiment, and forfeit all winnings (including the participation fee).

If you discover that you have put your guesses in the wrong place (say, the wrong page or wrong row), please raise your hand.

Draw	**Drawn Ball Colour**	**Your Probability Guess**
0		50%
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

Instructions for 0.6 Condition

These were identical to those for the 0.7 condition, except that ‘six balls’ (‘6 balls’) were replaced for ‘seven balls’ (‘7 balls’), and ‘four balls’ (‘4 balls’) for ‘three place’ (‘3 balls’).

APPENDIX D. COMPUTATION OF MEAN SUM OF SQUARES ERROR AND ROBUSTNESS ANALYSIS

The mean sum of squares error (MSE) is equal to the sum of squares error (SSE) divided by the number of relevant datapoints.

D.1 *Sign Tests*

For the purpose of nonparametric *sign tests*, an algorithm x fits better than an algorithm y to predict the choices by a given experimental participant if she has a lower MSE if we were to predict her choices using algorithm x than if we were to predict her choices using algorithm y. Therefore we are interested in computing MSE values *at the level of* each experimental participant, i.e. at the level of the goodness of fit of each algorithm for each experimental participant.

Define R the number of rounds that are included in the sample for each participant; $r = 15$ the number of rounds in a given period (always 15); q the number of periods that are included in the sample for each participant. Clearly, $R = rq$ by definition. There are three procedures to compute the MSE *at the level of* each experimental participant.

MSE by observation. It is possible to compute the mean squared difference by observation between prediction and observation. Then, in relation to each participant making a choice p_i^{actual} in round i and to each corresponding theoretical prediction p_i^{theory}, it is possible to compute:

$$MSE = \frac{SSE}{R} = \frac{\sum_{i=1}^{R}(p_i^{actual} - p_i^{theory})^2}{R}$$

where $R = rq$: $R = 90$ for the full sample and for most subjects in the restricted sample (i.e., whenever $q = 6$); $R = 75$ for the nine subjects with contaminated periods (one each) in the restricted sample (i.e., whenever $q = 5$); $R = 60$ for the experienced sample and for most subjects in the experienced restricted sample (i.e., whenever $q = 4$); $R = 45$ for the four subjects with contaminated periods (one each) among periods 3-6 in the experienced restricted sample (i.e., whenever $q = 3$).

MSE by observation values provide the natural measure in relation to which to estimate sign tests, as in section 5.4.

MSE by period. It is possible to compute the mean squared difference by period between predictions and observations. The relevant test statistic here is

$$MSE = \frac{SSE}{q} = \frac{\sum_{l=1}^{q}\left(\sum_{k=1}^{r} p_i^{actual} - \sum_{k=1}^{r} p_i^{theory}\right)^2}{q}$$

where $r = 15$ and the value of q depends on the number of periods in the sample: $q = 6$ in the full sample and in most cases in the restricted sample; $q = 5$ for the nine subjects with contaminated periods (one each) in the restricted sample; $q = 4$ in the experienced and in most cases in the experienced restricted sample; q = 3 for the four subjects with contaminated periods (one each) among periods 3-6 in the experienced restricted sample.

MSE by subject. It is possible to compute for each subject the square of the sum of differences between predictions and observations. For the purpose of sign tests where the analysis is at the level of goodness of fit at the level of each experimental participant, the MSE by subject is, by construction, the same as the SSE by subject. The relevant test statistic is

$$MSE = SSE = \left(\sum_{i=1}^{R} p_i^{actual} - \sum_{i=1}^{R} p_i^{theory} \right)^2$$

where $R = rq$, with possible values equal to 90, 75, 60 and 45, as specified earlier under the description of the MSE by observation algorithm.

D.2 *Robustness Tests*

In this appendix we run sign tests using MSE values computed by period or by subject rather than by observation as in the main text. The results are shown in Table 13.
(*Table 13 appears at the end of the document.*)
IE_N and IE_C always outperform RE. IE_N outperforms IE_C in the MSE analysis by period, and in two of the four samples in the MSE analysis by subject.

APPENDIX E. TESTS OF ROBUSTNESS OF α ESTIMATES

E.1 *Experimental Hypotheses*

This appendix discusses two additional hypotheses, designed to test the robustness of IE. The first hypothesis is a test of the robustness of α estimates to changes in the task they are estimated from.

HYPOTHESIS 3. Mean α values do not significantly differ between the 0.6 and 0.7 conditions.

We also aim to test whether there is some sense in which agents do *not* converge towards RE as the experiment progresses. To test this, we employ the least squares method to estimate α values that best fit each period as played by each subject. That is, we consider the 15 choices made by a given subject in a given period, and we find the *period-specific* value of α that mimimizes the sum of squared errors between predictions and observations. Thus, for each subject there are six period-specific values of α, and one can analyze whether these period-specific values followed any particular dynamic pattern in the experiment.

HYPOTHESIS 4. Mean period-specific α values tend to converge to 1 as the experiment progresses and the subjects have opportunities to learn about the nature of the task.

With Hypothesis 4, we aim to verify the absence of any obvious convergence towards greater RE play (i.e., IE with α = 1) across the 90 rounds of the experiment.

E.2 *Experimental Results*

Hypothesis 3. As shown by Table 3 in the main text, mean α values are surprisingly stable between the two conditions. While the size of the difference changes depending on the sample and IE measure used, Hypothesis 3 can never be rejected (at $P < 0.1$ or better) in Mann-Whitney tests. The greatest variability occurs for the full sample and for the restricted sample in relation to IE_N. In all other cases mean α values are within at most 0.06 of one another.

Hypothesis 4. The fact that the evidence for the other hypotheses is robust to the removal of the first two periods, as in the experienced sample, already suggests that IE may not tend to converge to RE as experience grows.

FIGURE 5. MEAN α ESTIMATES BY EXPERIMENTAL PERIOD

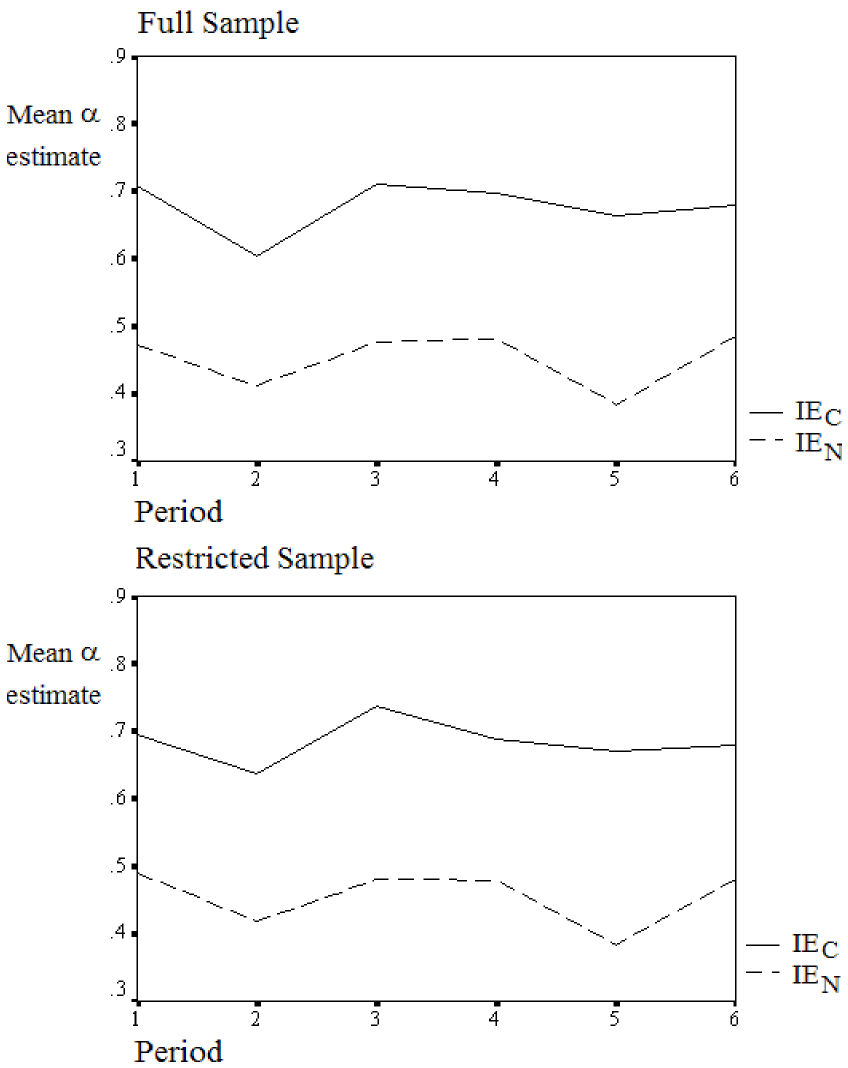


Figure 5 shows the absence of any convergence to RE across the ninety rounds of the experiment. Spearman correlation coefficients between mean α values and period number are not statistically significant (and are within 0.025 of 0) for all samples and both for IE_C and IE_N.

References

Akerlof, (2002) Behavioural Macroeconomics and Macroeconomic Behaviour, *American Economic Review* 92(3) June 2002, 411-433.

Beckman, S. R. & D. Downs (1997) Forecasters as Imperfect Information Processors: Experimental and Survey Evidence, *Journal of Economic Behavior and Organization* 32, 89-100.

Blomqvist, H. C. (1989) The 'Rationality' of Expectations: An Experimental Approach, *Journal of Economic Psychology* 10, 275-299.

Camerer, C. (1995) Individual Choice, in J.H. Kagel & A.E. Roth (eds.), *Handbook of Experimental Economics*, Princeton: Princeton University Press, 587-703.

Campbell, J. and R. Clarida (1987) The Dollar and Real Interest Rates, *Carnegie-Rochester Conference Series on Public Policy* 27, 103-140.

Davidson, R. & J. G. MacKinnon (1993) *Estimation and Inference in Econometrics*, New York and Oxford: Oxford University Press.

Davis, D. D. & C. A. Holt (1993) *Experimental Economics*, Princeton: Princeton University Press.

Dornbusch, R. (1976) Expectations and Exchange Rate Dynamics, *Journal of Political Economy* 84, 1161-1176.

Dwyer, G. P. Jr., Williams A. W., Battalio R. C. & T. I. Mason (1993) Tests of Rational Expectations in a Stark Setting, *Economic Journal* 103, 586-601.

Foster, D.P. & H. Peyton Young (2003) Learning, Hypothesis Testing, and Nash Equilibrium, *Games and Economic Behavior* 45 (1), 73-96.

Frankel, J. A. & K. A. Froot (1987) Using Survey Data to Test Standard Propositions Regarding Exchange Rate Expectations, *American Economic Review* 77, 133-153.

Frankel, J. A. & A. K. Rose (1995) Empirical Research on Nominal Exchange Rates, in G. Grossman & K. Rogoff, *Handbook of International Economics*, Amsterdam: Elsevier, 1689-1729.

Frydman, R. & M. D. Goldberg (2003) Imperfect Knowledge Expectations, Uncertainty-Adjusted Uncovered Interest Rate Parity, and Exchange Rate Dynamics, in P. Aghion, R. Frydman, J. Stiglitz & M. Woodford (eds.), *Knowledge, Information and Expectations in Modern Macroeconomics*, Princeton: Princeton University Press, 145-182.

Gerlach, S. & Smets, F. (1997), The term structure of Euro-rates: some evidence in support of the expectations hypothesis, *Journal of International Money and Finance* 16 (2), 305-321.

Gigerenzer, G., P. M. Todd & the ABC Research Group (1999) Simple Heuristics That Make Us Smart, Oxford: Oxford University Press.

Goldberg, M. D. & R. Frydman (1996), Imperfect Knowledge and Behavior in the Foreign Exchange market, *Economic Journal* 106, 869-893.

Gourinchas, P.-O. & A. Tornell (2004), Exchange Rate Puzzles and Distorted Beliefs, *Journal of International Economics* 64, 303-333.

Hall, R. (1978) Stochastic Implications of the Life Cycle Permanent Income Hypothesis: Theory and Evidence, *Journal of Political Economy* 85, 971-987.

Hamilton, J. (1989) A New Approach to the Economic Analysis of Non-stationary Time Series Subject to Changes in Regime, *Econometrica* 57, 357-384.

Kool, C. & D. Thornton (2004) A Note on the Expectations Hypothesis at the Founding of the Fed, *Journal of Banking and Finance* 28, 3055–3068.

Kuhn, T. S. (1970) *The Structure of Scientific Revolutions*, 2nd Ed., Chicago: University of Chicago Press.

Kydland, F. & E. C. Prescott (1982) Time to Build and Aggregate Fluctuations, *Econometrica* 50, 1345-1370.

Lucas, R. E., Jr. (1972) Expectations and the Neutrality of Money, *Journal of Economic Theory* 4, 103-124.

Maddock, R. & R. Carter (1982) A Child's Guide to Rational Expectations, *Journal of Economic Literature* 20, 39-51.

Mankiw, N. G. and Miron, A. J. (1986) The Changing Behavior of the Term Structure of Interest Rates, *Quarterly Journal of Economics* 101, 211-228.

Mankiw, N. G. & R. Reis (2002) Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve, *Quarterly Journal of Economics* 117, 1295-1328.

Mankiw, N. G. & R. Reis (2003) Sticky Information: A Model of Monetary Nonneutrality and Structural Slumps, in P. Aghion, R. Frydman, J. Stiglitz & M. Woodford (eds.), *Knowledge, Information and Expectations in Modern Macroeconomics*, Princeton: Princeton University Press, 64-86.

Menzies, G.D. & D.J. Zizzo (2004), Inferential Expectations, Oxford University Department of Economics Discussion Paper 187.

Rappoport, P. (1985) Unfalsified Expectations: An Alternative Perspective on Modelling Expectations in Macroeconomics, C.V. Starr Center for Applied Economics research report 85-16, New York University.

Rotheli, T. E. (1998) Pattern Recognition and Procedurally Rational Expectations, *Journal of Economic Behavior and Organization* 37, 71-90.

Schmalensee, R. (1976) An Experimental Study of Expectation Formation, *Econometrica* 44, 17-41.

Swenson, C.W. (1997) Rational Expectations and Tax Policy: Experimental Market Evidence, *Journal of Economic Behavior and Organization* 32, 433-455.

Woodford, M. (1991) Self-Fulfilling Expectations and Fluctuations in Aggregate Demand, in N. G. Mankiw & D. Romer (eds.), *New-Keynesian Economics*, vol. 2, Cambridge (MA): MIT Press, 77-110.

Zellner, A. (1988) Bayesian Analysis in Econometrics, *Journal of Econometrics* 37, 27-50.

TABLE 6 - GOODNESS OF FIT OF EXPECTATIONAL MODELS USING SIGN TESTS

		Full Sample				Experienced Sample			
x	y	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*
IE_C	IE_N	1	23	13	0.002	2	22	13	0.007
IE_C	RE	11	26	0	0.001	14	23	0	0
IE_N	RE	23	14	0	0	24	13	0	0
		Restricted Sample				Restricted Experienced Sample			
x	y	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*
IE_C	IE_N	1	23	13	0.002	0	23	14	0
IE_C	RE	11	26	0	0.001	13	22	2	0.007
IE_N	RE	23	14	0	0	24	13	0	0

Notes: Each cell value indicates the number of subjects for which, in relation to any two algorithms x and y (and a given sample), MSE(x) < MSE(y), or MSE(x) = MSE(y), or MSE(x) > MSE(y), where MSE is the mean sum of squares error by observation: see Appendix C. *P* values are approximated to three decimal places.

TABLE 13 - FURTHER SIGN TESTS OF THE GOODNESS OF FIT OF EXPECTATIONAL MODELS

A. TESTS FOR MSE BY PERIOD

		Full Sample				Experienced Sample			
x	y	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*
IE_C	IE_N	1	23	13	0.002	1	14	22	0.001
IE_C	RE	11	26	0	0.001	12	23	2	0.013
IE_N	RE	23	14	0	0	24	13	0	0
		Restricted Sample				Restricted Experienced Sample			
x	y	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*
IE_C	IE_N	1	23	13	0.002	1	14	22	0.001
IE_C	RE	11	26	0	0.001	12	23	2	0.013
IE_N	RE	23	14	0	0	24	13	0	0

Notes: MSE: mean sum of squares error *by period*. *P* values are approximated to three decimal places.

B. Tests for MSE by Subject

		Full Sample				Experienced Sample			
x	y	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*
IE_C	IE_N	7	23	7	1	2	22	13	0.007
IE_C	RE	10	26	1	0.012	11	23	3	0.057
IE_N	RE	17	14	6	0.035	20	13	4	0.002
		Restricted Sample				Restricted Experienced Sample			
x	y	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail *P*
IE_C	IE_N	7	23	7	1	2	22	13	0.007
IE_C	RE	10	26	1	0.012	11	23	3	0.057
IE_N	RE	17	14	6	0.035	20	13	4	0.002

MSE: mean sum of squares error *by subject*. *P* values are approximated to three decimal places.